Research Journal of Applied Sciences, Engineering and Technology 10(12): 1379-1383, 2015
DOI: 10.19026/rjaset.10.1837
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: March 19, 2015 Accepted: April 25, 2015 Published: August 25, 2015

Research Article

A Novel FPGA Based Low Cost Solution for Tamil-text to Speech Synthesizer

[1]T. Jayasankar and [2]J. Arputha Vijayaselvi
[1]Department of Electronics and Communication Engineering, Anna University BIT Campus, Tiruchirappalli, Tamilnadu 620024,
[2]Department of Electronics and Communication Engineering, Kings College of Engineering, Pudukkottai, Tamilnadu, India

Abstract: This study presents a prior work of developing a single chip solution for Text-to-Speech synthesizer for Tamil (Tamil-TTS) language. Though there are enormous works presented in the recent days to address TTS for their native languages, the motivation of this study is to develop a low-cost FPGA based solution for Tamil TTS synthesizer. This study uses the unique feature of Tamil language to eliminate the complexity involved in accessing a database of stored audio signals. It uses only the audio signals of consonants and vowels in the stored memory locations. The compound characters from the segmented input text are generated using a Direct Digital Synthesizer by operating at three different frequencies of phonetic interval units of Tamil. The proposed system is implemented in Cyclone IVE EP4CE115F29C7 FPGA device and the implementation results show that the proposed system outperforms the other similar methods in terms of memory utilization, text-to-speech time, area utilization and power dissipation. The accuracy of the system is examined with 25 native speakers and acceptable accuracy scale has been reached.

Keywords: FPGA implementation, speech synthesis, Tamil-TTS, text-to-speech

INTRODUCTION

In recent days, Text-to-Speech (TTS) is an attracting problem for many researchers to come with a catchy solution in both hardware and software. As the world is evolving around the internet and hand held devices, Text-To-Speech Synthesizer's has its own importance in par with other applications. Currently TTS is used in e-book reading, caller identification in mobiles, email reading services, news reading and announcement services (Violaro and Boeffard, 1998; Shih and Sproat, 1996; Klatt, 1987; Chazan *et al*., 2005). Most of these services tend to be in favor of visually and speech impaired users. These services are majorly implemented in hardware for English and very few researchers have implemented it for Mandarin (Shih and Sproat, 1996). These services are available in two different methods, one by a vast stored database access and the other by the phonetic pronunciation influenced by syllabification. In these two methods, the earlier approach has become outdated due to its higher memory utilization property, whereas the later is a promising technique which uses less memory by maintaining the accuracy level.

In both the techniques the speech synthesis process is organized as front end and back end process. In front end, the input text can be real time or stored data. Depending on the language, input text is processed into smaller elements (syllabification) and sent to the back end (Aida-Zade *et al*., 2010; Phan *et al*., 2014; Ferreira *et al*., 2014). Where these syllabified inputs are processed under speech related signal processing techniques at the back end. Though there is a vast need for developing better TTS, till date very few notable research works has been done for Indian regional languages (Hindi, Tamil, Telugu, etc.) (Rama *et al*., 2001; Sen and Samudravijaya, 2002; Bellur *et al*., 2011; Jayasankar and Vijayaselvi, 2014; Saraswathi and Vishalaksh, 2010; Sivaradje and Dananjayan, 2004).

The work presented here utilizes the unique feature of the regional language-Tamil, in which consonants (Vallinam-Mellinam-Idaiyinam) and vowels can be used to produce compound character sound. As said earlier, though TTS can be implemented in both hardware and software, very few researchers has taken the challenge of implementing TTS in hardware. The major hurdle in hardware implementation is the difficulty in accessing the stored vocabulary database. This study presents a novel TTS technique for Tamil language (TTS-Tamil) which operates only based on consonants and vowels stored in the database.

Speech synthesis-overview: The bottom line of Speech Synthesis process is the concatenation process. In

Corresponding Author: T. Jayasankar, Department of Electronics and Communication Engineering, Anna University BIT Campus, Tiruchirappalli, Tamilnadu 620024, India

Thirukkural-TTS by Rama *et al*. (2001), proposes both offline and process for Tamil Text-to-speech. The offline process combines five different stages, combining basic units, building database, study of prosody in natural speech, consonant-vowel segmentation and pitch marking. In the process of study of prosody includes the grammatical rules for proper pronunciation based on pauses and duration for the naturalness of the synthesized speech. These duration scales are stored in database as a look up table. When implemented this offline technique is capable of achieving 98% accuracy. On the other hand, in the online process the process of building database is eliminated for sampling the synthesizing process. These both methods seem to be a prominent solution for TTS (any native language) for a software implementation. Though it is mentioned that, these methods are prone to low distortion, it is not prudent to use this scheme for a hardware implementation. Similarly, TTS for web browsing by Sen and Samudravijaya (2002), proposes a online solution to get rid of the memory issues for storing a database. This scheme is developed for both Hindi and English text contents and uses exhaustive rule sets. But, according to the author's statement, the naturalness of the synthesized speech has to be improved. Bellur *et al*. (2011), developed a prosody TTS model for Hindi and Tamil in which Classification and Regression Trees (CART) was modified to syllable-based synthesis. The Mean opinion score for the system gets a scale greater than 3, which is a nominal score rated from 1 to 5. In the recent scenario, came up with a syllable based TTS scheme for Tamil which uses neural network for prosody prediction. This concatenative speech synthesis scheme uses five layers of auto associative neural network to get better naturalness to the final processed speech signal. It is proved that, the TTS with prosody has better naturalness than the TTS scheme without prosody.

In all these techniques, the process of achieving the naturalness (i.e., accuracy) majorly depends on the prosody prediction techniques. Some of the methods studied here present neural and fuzzy logic concepts for the synthesis process which will be more complex when implemented in hardware. Addressing the above said issue (for Tamil-TTS), Sivaradje and Dananjayan (2004), designed and implemented TTS converter for satellite radio receivers for FPGA. A much comprehensible work has been done by Jayasankar and Vijayaselvi (2014), to examine the real time difficulties in implementing TTS-Tamil in FPGA. It follows a set of condensed rules for segmentation of tamil words. As the implementation part is done in Verilog, the input text in given in English with the Tamil pronunciation. Further, this study presents a novel technique with a speech synthesis technique which is implemented in FPGA.

DESIGN OF SPEECH SYNTHESIZING UNIT

The design process of the novel Speech Synthesizing technique is developed with much care considering memory utilization issues raised in the other FPGA based TTS methods (Sivaradje and Dananjayan, 2004; Bamini, 2003; Khalifa *et al*., 2008). One major drawback in both software and hardware designs is the memory utilized for storing the words either offline or online memory devices. Although certain TTS schemes considers memory utilization (database) as a trade-off parameter and concentrates on 100% accuracy, the objective of our work is to develop a FPGA based low-cost standalone TTS scheme for which we have studied the unique features of Tamil language and combined it to develop the novel TTS system (Fig. 1).

When the need for low cost solutions were demanded, many researchers came up optimized solutions like syllabification based on prosody predictions. Such techniques can provide optimized results for languages like English which has only 26 characters. But, developing such system for a language like Tamil which has 247 characters will be a quiet difficult task.

To reduce this complexity we utilize the feature of producing compound character from consonants (Vallinam-Mellinam-Idaiyinam) and vowels (Uyirezhuthu). The pronunciation accuracy of these characters depends of the time slot taken to spell out each the characters which is measured in terms of Mathirai (unit of phonetic interval).

An indigenous Direct Digital Synthesizer generates three variant frequencies for three different units of phonetic interval. Mathirai units of consonants, kuril and nedil characters are half second, one second and two seconds respectively. This synthesize unit concatenates these compound signals based on the phonetic intervals and generates the speech signal at the receiver end. The design of the proposed system is shown in Fig. 2. Further, the implementation process is explained in the next section.

Implementation of speech synthesizing unit: In this chapter, the design of novel Tamil-TTS technique in a standalone FPGA device is explained in detail. The overall schematic of the proposed system is shown in Fig. 2. SD-Card contains the audio samples of Consonants and Vowel characters for the synthesizing process. During the initialization process, our proposed TTS-Core fetches the Audio Signal (s) from SD-Card and stores in on-board SRAM memory.

The need for this process has risen to achieve a faster TTS scheme than conventional methods. The TTS-Core and text analyzer are interfaced with NIOS-core through Avalon Bus. In the preprocessing stage, the input text read from the PS2 controller is sent to the text analyzer, where the input data stream will be segmented into compound character (Consonants+ Vowels). In parallel, text analyzer estimates the

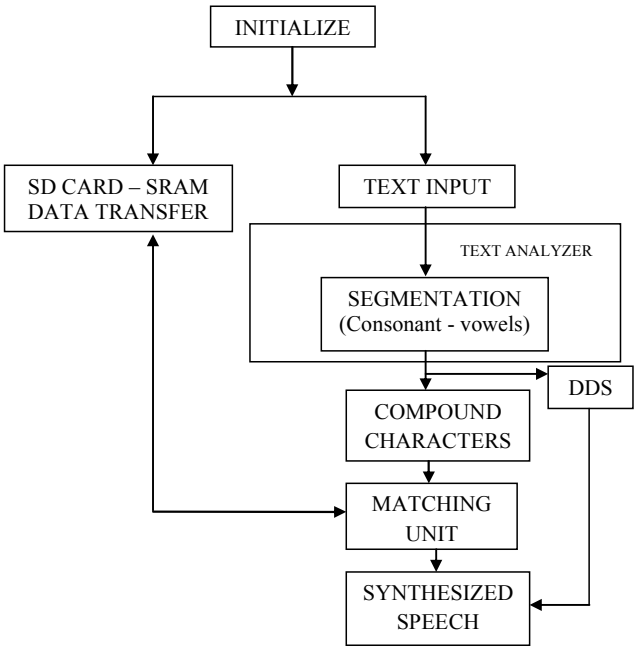

Fig. 1: Flow diagram of the proposed Tamil-TTS

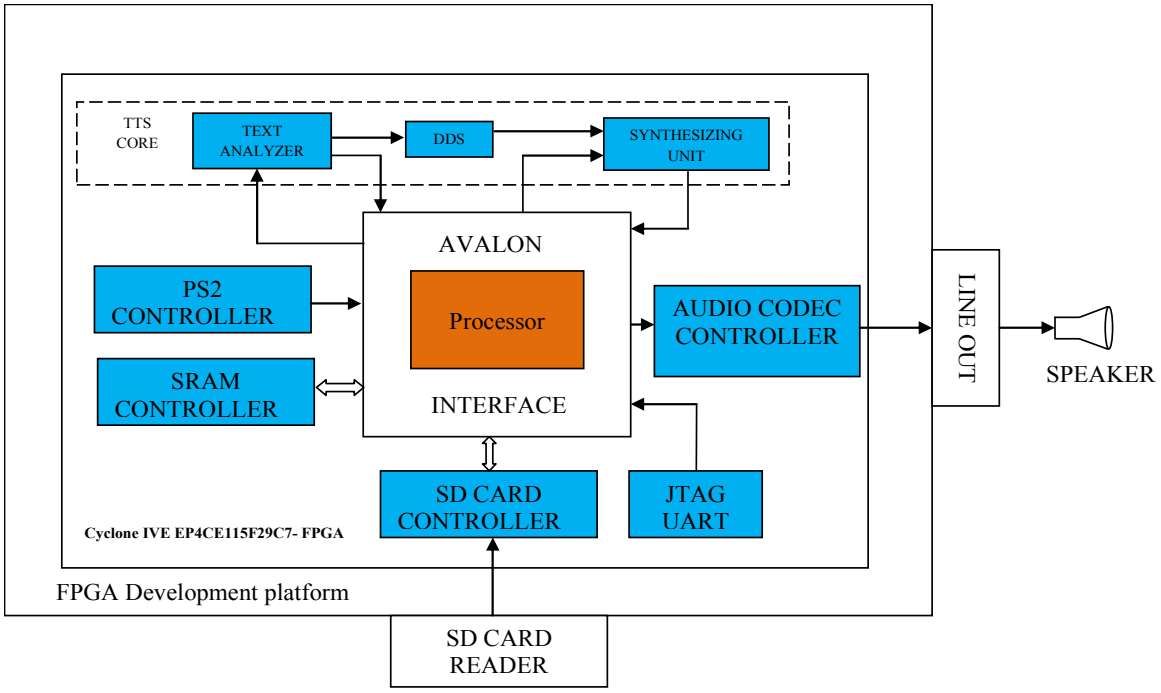

Fig. 2: FPGA schematic of the proposed Tamil-TTS scheme

Mathirai unit (unit of phonetic interval) for each segmented characters.

Based on this phonetic interval unit, DDS estimates frequency variation for the concatenation process. Corresponding audio signal (s) for the consonants and vowels are put in frequency matching stage and the concatenated compound character signals are converted using DAC by Audio Codec Controller for the final speech output. The pseudo code of TTS as shown in Fig. 3.

RESULTS AND DISCUSSION

Voice quality testing is performed using subjective test. In subjective tests, human listeners hear and rank the quality of processed voice files according to a

Parameter define-source: Start_Address
Destination: Start_Address
// read a block of data from SD-Card and Store in SRAM (512 byte)
enable CMD17
read_en = 1'b1
write_en = 1'b1
@ pos clock
…
If (!end_of_data)
SRAM (Start_Address) <= SD (Start_Address)
end
	//synchronize PS2 with the system
	@ system clock 50 MHz
	@ pos (ps2_clock && reset)
	..
	Rx_data = 8 bit key in word
	Sram_buff <= Rx_data
	..
	when new data is received
	Rx_data_en = 1'b1
	repeat Read operation
	..
	end of PS2 read operation

//text analyzer
read_sram_en = 1'b1
segment (word)
if (current char_val = vowels)
compound_char = buf (current_add: start_add)
DDS (compound_char) //compound_char <= cons+vow
End return: compound_char, mathirai_unit
	//matching unit
	enable row && column scan
	define Audio enable delimiter
	match (compound_char) in SRAM (start_address: end_address)
	end of scan
	return: DDS (compound_char, match_address)

//DDS
Parameter Define: $freq_{half}$, $freq_{one}$, $freq_{two}$
If (mathirai_unit (compound_char [7:0]) == 2'b01)
set compound_char_cons_freq = $freq_{half}$
end
If (mathirai_unit (compound_char [15:0]) == 2'b10)
set compound_char_vowelk_freq = $freq_{one}$
end
If (mathirai_unit (compound_char [15:0]) == 2'b11)
set compound_char_vowleln_freq = $freq_{two}$
end
..
@ pos clock
for (current_scan_word)
audcod (match_address of all compound_char of current scan word)
end	

Fig. 3: Pseudo code for TTS

Table 1: Evaluation results

Database content	Memory	Power dissipation (mW)	Accuracy Min. score (1) Max. score (5)	Logic elements	Time-TTS (msec)
Consonant+vowel (proposed)	30	98.93	4.27	373	12
Compound character	300	87.03	4.14	1744	19
Words	66	57.34	5.00	154	72

Min.: Minimum; Max.: Maximum

Appendix A: classification of Tamil letters

Consonants	Vallinam	க், ச், ட், த், ப், ற்
	Mellinam	ங், ஞ், ண், ந், ம், ன்
	Idaiyinam	ய், ர், ல், வ், ழ், ள்
Vowels	Kuril	அ, இ, உ, எ, ஒ
	Nedil	ஆ, ஈ, ஊ, ஏ, ஐ, ஓ, ஔ
	Ottru	ஃ

Appendix B: Vocabulary set of 16 words

Single letter word	Two letter word	Three letter word	Four letter word
கை, தை,	படை, பாறை, தடை, தொகை கறை, கடை, கதை	பற்று, பகடை, படகு, காற்று, தச்சு	கசப்பு

certain scale. The most common scale is called a Mean Opinion Score (MOS) and is composed of 5 scores of subjective quality, 1-Bad, 2-Poor, 3-Fair, 4-Good, 5-Excellent. The MOS score of a certain TTS system is the average of all the ranks voted by different listeners of the different voice file used in the experiment. The tests were conducted in a laboratory environment with 25 students in the age group of 20-28 years by playing the synthesized Tamil speech signals through headphones. In this case, the subjects should possess

the adequate speech knowledge for accurate assessment of the speech signals. The performance and quality assessment of the proposed system has been evaluated with different cases of database contents as shown in Table 1. For the proposes system, audio signals of consonants and vowels listed in Appendix A are stored in SD-Card. As explained in the Design of Speech Synthesizing unit Section, during the initialization process of TTS-Scheme, memory controller copies the signals stored in the SD-Card to S-RAM through the Avalon Interface in the Nios Processor. On the second case audio signals of compound characters are stored. For the third case, a vocabulary set of 16 words listed in Appendix B are stored. In all these three scenarios, the overall QoS of the proposed method is in the acceptable scale. Though the accuracy of the proposed is less than the accuracy achieved through the vocabulary set, it is quite higher than the second case. The highlighting QoS parameter is Time-TTS (Time taken for Text-to-Speech).

The proposed scheme outperforms all the other two cases with an average Time-TTS as 12 msec. As the proposed system uses lesser memory allocation for the audio signals, on-chip memory utilization can be considered as a future extension of this current study.

CONCLUSION

The demand for developing a Text-to-Speech synthesizer for Tamil language has been addressed and solved in this study. The standalone FPGA based TTS synthesizer uses the unique features of the native Tamil language to reduce the memory complexity issues in hardware implementation. Hence, much attention has been put forth developing a TTS-Core, with Direct Digital Synthesizer to produce a satisfactory speech signal with less area utilization and much lesser time for processing the speech output. We conclude from this study the real time implementation results show that the proposed Tamil-TTS with the stored vowel and consonant sounds requires 90% lesser memory than the conventional techniques and easy method of synthesizing speech for Tamil Language. As a future work, this study can be extended by using a on-chip memories in FPGAs to get a faster and cheaper solution for Tamil-TTS.

REFERENCES

Aida-Zade, K.R., C. Ardil and A.M. Sharifova, 2010. The main principles of text-to-speech synthesis system. Int. J. Comput. Electr. Automat. Control Inform. Eng., 7(3): 234-240.

Bamini, P.K., 2003. FPGA-based implementation of concatenative speech synthesis algorithm. M.Sc. Thesis, University of South Florida.

Bellur, A., K.B. Narayan, K.R. Krishnan and H.A. Murthy, 2011. Prosody modeling for syllable-based concatenative speech synthesis of Hindi and Tamil. Proceeding of the National Conference on Communications (NCC). Bangalore, pp: 1-5.

Chazan, D., R. Hoory, Z. Kons, A. Sagi, S. Shechtman and A. Sorin, 2005. Small footprint concatenative text-to-speech synthesis system using complex spectral envelope modeling. Proceeding of the Interspeech. Lisbon, Portugal, pp: 2569-2572.

Ferreira, J.P., C. Chesi, H. Cho, D. Baldewijns, D. Braga *et al*., 2014. On mirandese language resources for text-to-speech. Proceeding of the 4th International Workshop on Spoken Language Technologies for Under-resourced Languages.

Jayasankar, T. and J.A. Vijayaselvi, 2014. FPGA-based implementation of text analyser and syllable preparation for concatenative speech synthesis of tamil language. Aust. J. Basic Appl. Sci., 8(10): 102-109.

Khalifa, O.O., Z.H. Ahmad, A.H.A. Hashim and T.S. Gunawan, 2008. SMaTalk: Standard malay text to speech talk system. Signal Process. Int. J., 2(5): 1-16.

Klatt, D.H., 1987. Review of text-to-speech conversion for english. J. Acoust. Soc. Am., 82(3): 737-793.

Phan, T.S., A.T. Dinh, T.T. Vu and C.M. Luong, 2014. An Improvement of Prosodic Characteristics in Vietnamese Text to Speech System. In: Huynh, V.N. *et al*. (Eds.), Knowledge and System Engineering. Advances in Intelligent Systems and Computing, Springer International Publishing Switzerland, 244: 99-111.

Rama, G.L.J., A.G. Ramakrishnan, R. Muralishankar and V. Venkatesh, 2001. Thirukkural-a text to speech synthesis system. Proceeding of the Tamil Internet, pp: 92-97.

Saraswathi, S. and R. Vishalaksh, 2010. Design of multilingual speech synthesis system. Int. J. Intell. Inform. Manage., 2: 58-64.

Sen, A. and K. Samudravijaya, 2002. Indian accent text-to-speech system for web browsing. Sadhana, 27(1): 113-126.

Shih, C. and R. Sproat, 1996. Issues in text-to-speech conversion for mandarin. Comput. Linguist. Chinese Lang. Process., 1(1): 37-86.

Sivaradje, G. and P. Dananjayan, 2004. VHDL implementation of text to speech converter for satellite radio receivers. Proceeding of the National Communication Conference (NCC, 2004). Bangalore, India.

Violaro, F. and O. Boeffard, 1998. A hybrid model for text-to-speech synthesis. IEEE T. Speech Audi. P., 6(5): 426-434.