A SURVEY OF CONCATENATIVE TAMIL SPEECH SYNTHESIZER METHODS IN NATURAL LANGUAGE PROCESSING

Dr.J.Indumathi[1], M.Sharmila[2]
[1]indumathi@annauniv.edu, [2]sharmi.m.k@gmail.com

Department of Information Science and Technology
College of Engineering, Guindy,
Anna University, Chennai-25, Tamilnadu, India.

Abstract - *Natural language processing is the emerging approach that processing and analyzing the text using technologies. Recent research perspective in natural language processing are machine learning, speech synthesizer, voice recognition, spellchecker etc. The speech synthesizer technologies are day to day improved on both hardware and software platforms. The hardware relates to the speech processing with design of processor and chip. The synthesizer software analyze the text and interact with users. The speech synthesized methods are broadly divided into Formant, Concatenative and Articulatory synthesis. Formant synthesizer uses an acoustic model of the speech. This is used for mobile computing and embedded systems. Concatenative method is used in modern text to speech engines and produces natural sounding speech. Articulatory synthesis also uses acoustic model and produces understandable speech. The issues in these methods based on deriving rules, distortion, memory requirements, collecting the samples and quality results. The issues of Concatenative speech are alignment of recorded speech, automatic segmentation, optimized design, unit selection and automatic segmentation. This paper describes the Concatenative speech synthesis method for Tamil language with detailed survey and conquer the issues. The applications mainly focused on deafened and handicapped people, multimedia, communication field and all type of human-machine interactions.*

Keywords: Natural language Processing, Polysyllable, Harmonic plus Noise Model (HNM), Linear Predictive code (LPC), Text To Speech (TTS), linguistics

1. INTRODUCTION

Recent trends in digital communication techniques are increased in a wide manner. These techniques used for speech synthesis methods. Tamil is the official language of South India. Tamil language contains 2500 phonemes. In Tamil language, the consonants and vowels are 18 and 12 respectively. The handicapped people, multimedia and communication users are needed to get synthesized speech instead of distorted natural sound signal. The Text To Speech (TTS) synthesis process the textual input to speech. The main requirements for text to speech synthesizer are intelligible and natural outcome. The speech synthesis model mainly represent as text to speech synthesis, text processing, phonetic analysis, Prosodic analysis, prosodic modeling. The text to speech synthesis is to translate the random input text to natural sound. It utilizes linguistic analysis to identify the correct pronunciations and prosody and outcome as auditory sound. TTS is comprised into two components, one is natural language processing that is input text synthesis and another is digital signal processing that is output speech synthesis [Aimilios.et.al, (2010)].

The natural language processing involves the conversion of text into linguistic representation and in output this representation is converted into sound. The text processing is liable to find out all knowledge about the text. The role of text processing are linguistic analysis, document structure detection, text normalization and markup interpretation. The phonetic analysis use is to tag each word in the text and analyze the outcome sound. The job is to analyze morphs, grapheme and homograph disambiguation.

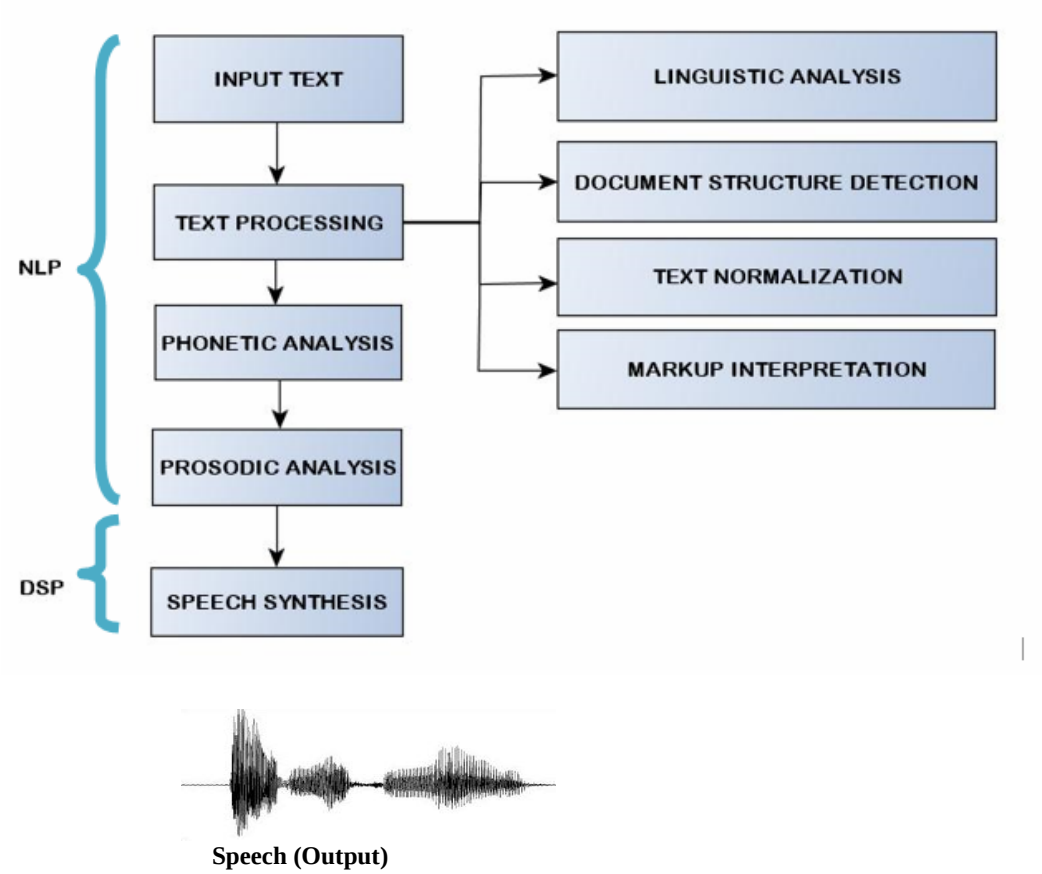

Figure 1. TEXT TO SPEECH FOR TAMIL

The prosodic analysis for the output level. This analysis controls the amplitude, duration and pitch of the sound. The approaches of prosodic models are rule based approach, statistical approach, as-is approach, Klatt's duration model, CART-based model, Neural network-based model, sum of products, Pierrehumbert's intonation model, tilt model and Fujisaki's intonation model. The figure.1 shows the text to speech for tamil. The synthesis techniques are Formant synthesis, Articulatory synthesis, Concatenative synthesis, unit selection synthesis, Hidden Markov Model-based synthesis and Harmonic plus Noise Model (HNM).

2. FORMANT SPEECH SYNTHESIS

In Formant Speech synthesis the process of formant resonance frequencies and amplitudes for vocal cavity. Stimulating a set of resonators in the source input and to obtain the speech signal. Five formants are needed to produce intelligible sound. The rule based formant synthesizer is based on set of rules. The cascade format synthesizer is also used for speech synthesis, in which the formants are connected in series. Similarly, the parallel formant synthesizer in which the formants are connected in parallel. The issues in formant speech synthesis are the samples are not used in runtime, provide intelligible sound but not natural, less memory in the resonators and useful for limited devices [Sukanya.et.al (2008)].

3. ARTICULATORY SYNTHESIS

Articulatory speech synthesis models the frequency of human articular behaviour. Practically, it is difficult to implement. The control parameters used are tongue tip position, tongue height, etc,. The issues are obtaining data for this modeling, hard to balance high level and low level models. The outcome of speech is not natural [Madiha.et.al, (2011)]. Both formant and articulatory synthesis has the limitation as difficult to get the output parameters from input text analysis.

4. CONCATENATIVE SPEECH SYNTHESIS

The basic input text is either of phones, diphones, triphones, polysyllables and syllables. The phones are inefficient for signal processing unit. The quantity of phone units in Tamil language is less than 50. The record of phones is small, as a result the dynamics of speech sound with large changeable is not achieved. The tamil diphone units are 1000 to 2000. The diphone concatenation generates natural speech based on prosody rules. A triphone is diphone added with one unit. The tamil triphones recorded are large in number compared to diphones and phones. They are adjusted with previous and next phones in a given phrase. Tamil languages are mostly based on syllables. Using syllable is the basic unit, the result produced are intelligible speech sound. Polysyllable is a basic unit which picks up trisyllable, followed by bisyllable and monosyllable units [Karunesh Arora (2013)]. It selects the largest unit in the database. Polysyllable provides the best quality speech in concatenative speech synthesis. The clustering of the syllable is based on consonants (C) and vowels (V). The general design is C*VC* where C* denotes the presence of 0 or more consonants. The clustering in Tamil language syllables is analyzed and improve the quality of speech [Tamar.et.al, (2012)].

The Corpus-Based Concatenative Speech Synthesis System for Turkish uses text corpus, speech corpus and unit selection process. The text corpus collects all the information (syllables, phrases and corpus size). The speech corpus data indicates the major effect of speech quality. The linguistic process is depends on Turkish pronunciation lexicon, phoneme conversion and prosodic analysis. The unit selection using Viterbi algorithm was proposed [Hasim SAK. et.al, (2006)]. Uniform Concatenative Excitation Model for Synthesising Speech without Voiced/Unvoiced Classification proposed an excitation model which can synthesize both voiced and unvoiced [Joao P. Cabral (2013)]. The LPC vocoder evaluation was performed by pitch-tracking algorithm. Articulatory-based version in a concatenative speech synthesizer was proposed and to overcome the problem of automatically generating utterances [Tao.et.al, 2012)]. [Karunesh et. al (2013)] discussed concatenative text to speech synthesis for hindi. In this paper, the text is recorded in anechoic chamber and the prosody prediction identifies the energy and pitch.[Ouni et. al (2013)] proposed bimodal acoustic-visual synthesis, the result is 3D face animation with acoustic speech.

4.1 PROBLEM OF CONCATENATIVE TAMIL SPEECH SYNTHESIS

The issues of concatenative tamil speech synthesis are alignment of recorded speech, optimized design, unit selection and automatic segmentation. The concatenative method mainly works on the basis of augmenting the pre-recorded speech. Time domain features and spectral domain variance are the major role in automatic segmentation of concatenation phase. The large problems occured in alignment and segmentation. The diphone concatenation uses only two units, optimized design is not possible. The design of automatic segmentation and alignment will provide the result of intelligible Tamil speech signal.

4.2 SOLUTION OF CONCATENATIVE TAMIL SPEECH SYNTHESIS

One of the solution for the issues mentioned, the Tamil phoneme unit chosen as polysyllable. The input Tamil text is confer to the natural language processing. The text normalization

analysis tokenization, lexical access and morphological analysis and Grammatical analysis. Polysyllable comprises of bisyllable, trisyllable and monosyllable units. The clustering of the polysyllables with prosodic analysis will produce the speech with accurate pitch, energy and duration. The result of natural language processing is given to the digital signal processing input. The concatenation unit synthesize the speech signal. The Linear Predictive code (LPC) synthesizer will produce the intelligible, quality and natural sound. The concatenation is followed by LPC.

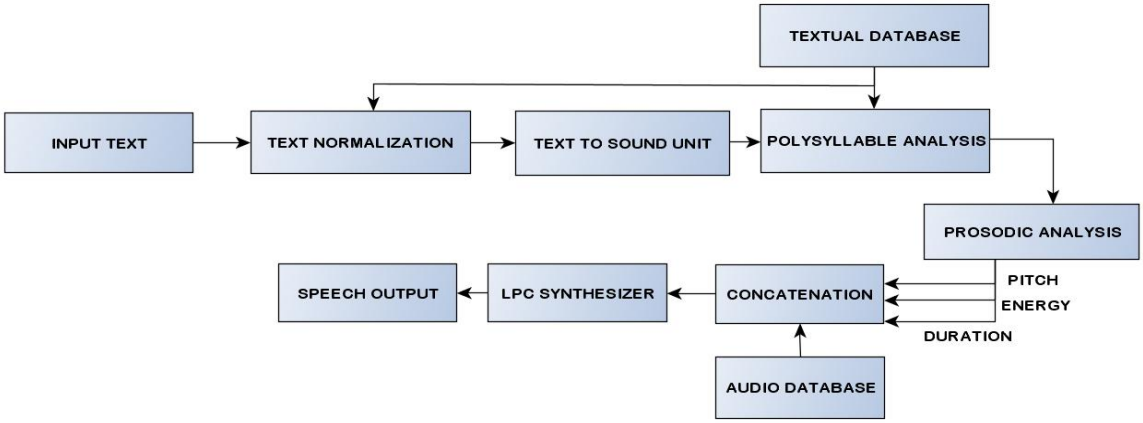

Figure 2. CONCATENATIVE TAMIL TEXT TO SPEECH SYNTHESIZER

Spectral mismatches is overcome by combining two units and each unit the distortion measurement is calculated by Euclidean distance. The pitch, energy, frame synchronous and duration of each units are merged. The merging process is in terms of variable length unit. The selection of variable length unit deals the different consonants and vowels for text to speech synthesis. The figure.2 shows the concatenation process for Tamil polysyllable analysis.

5. CONCLUSION

Tamil text to speech system composed of phonemes, consonants and vowels. This paper presented the main aspects of various text to speech system and issues in concatenative speech synthesizer. The concatenative text to speech synthesis method is well suited for Tamil language structure. Selection of the polysyllable speech unit and adjoin the prosodic information play a essential role in the development of concatenative tamil text to speech synthesis. The variable length concatenation provides a natural speech signal. The quality of speech in terms of cost and time is the challenge one for using polysyllabic concatenation. Another main challenge is in LPC synthesis, frame of the signal is dependent on the previous frames. These challenges affect the synthesized result. Future evaluations are the speed estimation of speech quality and triumph over the frame overlaps.

6. REFERENCES

- **[Aimilios.et.al, (2010)]** Aimilios Chalamandaris, Sotiris Karabetsos, Pirros Tsiakoulis, and Spyros Raptis, " A Unit Selection Text-to-Speech Synthesis System Optimized for Use with Screen Readers" *IEEE Transactions on Consumer Electronics*, Vol. 56, No. 3, pp.1890-1897, August 2010.

- **[Ashwin.et.al, (2011)]** Ashwin Bellur, K Badri Narayan, Raghava Krishnan K, Hema A Murthy, “Prosody Modeling for Syllable-Based Concatenative Speech Synthesis of Hindi and Tamil”, *Proc.2011 IEEE Conf.* 10.1109/NCC.2011.5734737.

- **[Hasim sak. et.al, (2006)]** Hasim sak, Tunga gungor, Yasar safkan, "A Corpus-Based Concatenative Speech SynthesisSystem for Turkish ", *Turk J Elec Engin*, Vol.14, No.2 2006

- **[Joao P. Cabral (2013)]** Joao P. Cabral Uniform Concatenative Excitation Model for Synthesising Speech without Voiced/Unvoiced Classification presented at *INTERSPEECH*, ISCA 2013 pp. 1082-1086.

- **[Juergen (2011)]** Juergen Schroeter AT & T labs-Research, Speech Signal Processing Guest Lecture at Boston university 2011 seminars, Available on http://www.bu.edu/dbin/hrc/ calendar/archive.php?y=2011

- **[Karunesh et. al (2013)]** Karunesh Arora, Sunita Arora, Mukund Kumar Roy "Speech to speech translation: a communication boon" *Springer CSIT*1(3):207–213, Sep 2013

- **[Madiha.et.al, (2011)]** Madiha J alil, Faran Awais Butt, Ahmed Malik, "A Survey of Different Speech Synthesis Techniques"*Proc.2013 IEEE Conf.*ISBN: 978-1-4673-5613-8/.

- **[Ouni et. al (2013)]** Slim Ouni, Vincent Colotte, Utpala Musti, Asterios Toutios, Brigitte Wrobel-Dautcourt, Marie-Odile Berger and Caroline Lavecchia, "Acoustic-visual synthesis technique using bimodal unit-selection" *EURASIP Journal on Audio, Speech, and Music Processing 2013*, Available on http://asmp.eurasipjournals.com/content/2013/1/16

- **[Sukanya.et.al (2008)]** Sukanya Yimngam, Wichian Premchaisawadi, Worapoj Kreesuradej " State of the Art Review on Thai Text-to-Speech System" presented at International Conference on Computer Science and Information Technology ISBN : 978-0-7695-3308-7/2008.
- **[Tao.et.al, 2012)]** Tao Jiang, Zhiyong Wu, Jia Jia, Lianhong Cai, "Perceptual Clustering Based Unit Selection Optimization ForConcatenative Text-To-Speech Synthesis" presented at ISCSLP 2012, ISBN.978-1-4673-2507-3/12.

- **[T.Jayasankar.et.al, (2011)]** T.Jayasankar, R.Thangarajan, J.Arputha Vijaya Selvi, “Automatic Continuous Speech Segmentation to Improve Tamil Text-to-Speech Synthesis”, International *Journal of Computer Applications*, Volume 25– No.1, 0975 – 8887, July 2011.

- **[Tamar.et.al, (2012)]** Tamar Shoham, David Malahand Slava Shechtman, "Quality Preserving Compression of a Concatenative Text-To-Speech Acoustic Database" *IEEE Transactions On Audio, Speech, And Language Processing*, Vol. 20, No. 3, March 2012.