Volume 5, Issue 2, February 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A System on Text To Speech Synthesis and Different Lilts in Speech

[1]Reshma*, [2]Chetna, [3]Amarbir Singh

[1, 2] Department of Computer Science, Chandigarh University, Gharuan, Mohali, Punjab, India

[3]Assistant Professor, Department of Mechanical Engineering, Chandigarh University, Gharuan, Mohali, Punjab, India

Abstract— ***Speech is the most natural way to communicate with each other. As per the requirement of advanced educational and information system, speech also increases the importance of human computer interaction. This paper presents an overview of Text To Speech Synthesis which converts text into the speech and describes the three types of voice lilts.***

Keywords— ***Synthesis, NLP, Phonetic Analysis, Lilts, Pitch***

I. INTRODUCTION

The objective for Text to Speech Synthesis is to convert the input text into the speech waveform. This task is categorized in two main fields text processing which is a part of Natural Language Processing and speech processing which is a part of Digital Signal Processing. Text Processing comprises of Text Analysis, Text Normalization and Phonetic Analysis. Text Analysis is initial task while performing Text to speech synthesis which analyze the input text and convert it into manageable form. Input Text consist of numbers, abbreviations, acronyms and idiomatic and transforms them into full text when needed. Text Normalization is used for generating synthesized speech. It transforms text into pronounceable form. The objective for this process to identify pauses and punctuation marks. Phonetic Analysis converts grapheme into phonemes. Phone is the smallest unit of sound and phoneme is collection of phones. Grapheme is collection of such various phonemes. Speech processing comprises of Prosodic Modeling and Acoustic Processing. Prosodic modeling describes speaker’s emotion and acoustic processing describes speaker’s characteristics.

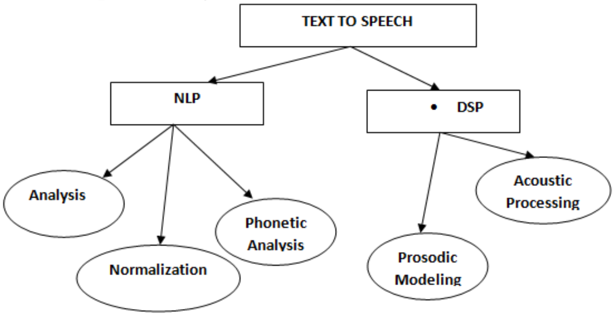


Figure 1 Text To speech Synthesis

II. NEED FOR TTS APPLICATION

Speech Synthesis is an important abetment tool and it has wide range application surrounding us. It provides an environment a person with disabilities can communicate. It’s important application is the screen readers for the person facing visual problems. It helps visually challenged person to continue their education and helps them to become a part of society. It is a multimedia package that provides facility of converting text into speech.

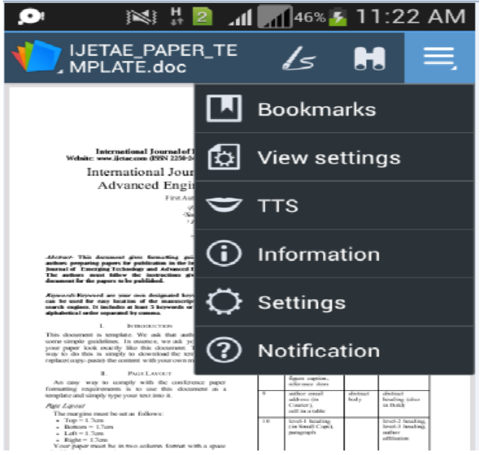
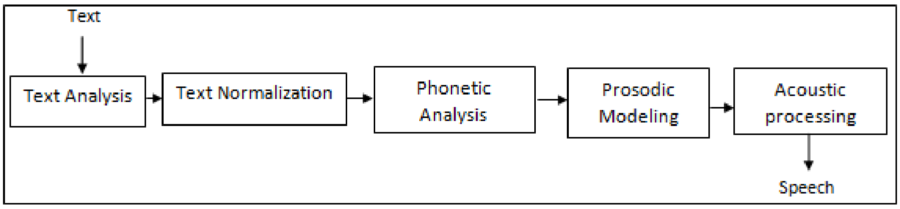


Figure 2 Inbuilt TTS by Google

III. METHODOLOGY



A. Text Analysis

The text analysis part is pre-processing part which analyzes the input text and organizes into manageable form. It consists of numbers, acronyms, idiomatic and abbreviations and converts them into full text when needed.

B. Text Normalization

Text Normalization makes the text into pronounceable form. The main objective for this process is to identify punctuation marks and pauses between words.

Number Converter: Pronunciation of numbers is different in different Scenario e.g. 1991 is pronounced as Nineteen Ninety One in some case and One Nine Nine One in another case or One Thousand Nine Hundred Ninety One.

*Acronyms Converter***:** Acronyms are converted into full textual form e.g. Er. Is pronounced as Engineer and Dr. are pronounced as Doctor and same like for other abbreviations.

*Word Segmentation***:** Collection of tokens make the sentence and token can be any word, acronym, abbreviation ,punctuation and delimiter e.g. “AICTE “- Aye eye see tee ee and “NCERT”- an see ee are tee. Token like Roman words also converted as iii – three etc.

C. Phonetic Analysis

Phone is smallest unit of sound and phoneme is collection of those phones. Grapheme is groupism of phoneme. Phonetic Analysis converts grapheme into phonemes.

D. Prosodic Modeling

Prosodic modeling describes the pitch of the speaker that how human being speaks a particular word. It generates prosodic properties of speech. It manages the pauses, durations in the speech. Variation in speech categorized by three ways-Slow Lilts, Fast Lilt

E. Acoustic Processing

Acoustic processing is the study of voice properties of a person while speaking. It includes the mechanical waves generated during speech and it is divided into three categories**.**

Concatenative Synthesis: As clear from the name Concatenative Synthesis is succession of recorded voices. So for maintaining the large set of voices we have to create and maintain the large database.

*Formant synthesis***:** This synthesis is more comprehensible than concatenative synthesis as there is no need to create a large database because the speeches are generated artificially and robotically.

*Articulatory Synthesis***:** This is synthesis of speech which uses mathematical model or techniques based on the replica of the human vocal tract and the processes occurring there. Vocal tract includes tongue, jaw and lips and speech is generated when the air passes through the vocal tract or by the different positions of the various speech articulators.

IV. RESULTS

A. SLOW LILT

In this style of pitch frequency of wave is very small. It almost complete one cycle or less in a second.

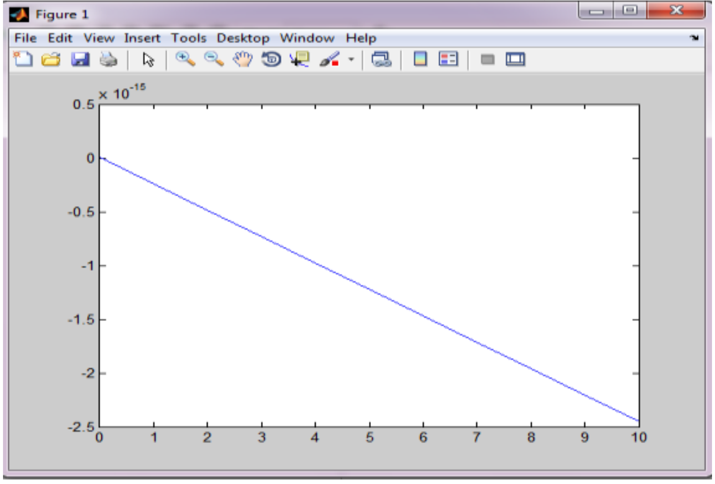


Figure 4.1 Slow Lilt

B. NORMAL LILT

In this Style of pitch frequency of wave is neither slow nor fast. It almost completes 10 cycles or more per second.

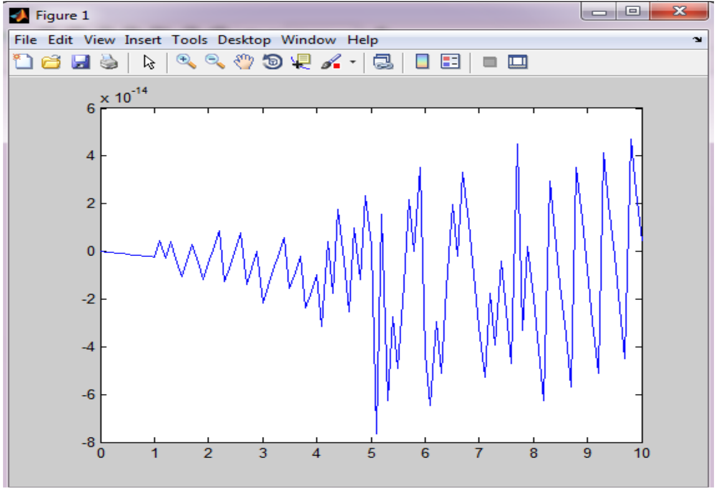


Figure 4.2 Normal Lilt

C.FAST LILT

In this style of pitch frequency of wave is very large. It almost complete 50-100 cycles per second.

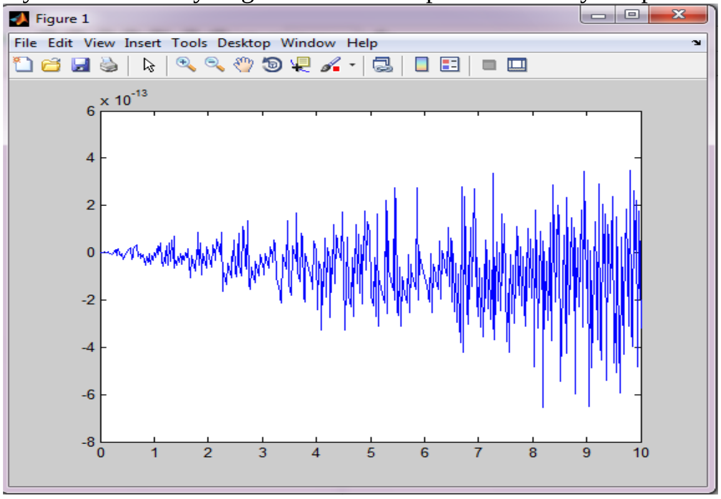


Figure 4.3 Fast Lilt

Table 1 Comparison of various lilts

Sr. No.	**Type of Lilt**	**Rate**	**Frequency**
1	Slow Lilt	-10	1
2	Normal Lilt	1-2	10
3	Fast Lilt	10	50-100

V. CONCLUSION

This paper describes victorious implementation of text to speech synthesis by using Matlab inbuilt functions of speech and enhance speech with different lilts. In future emotions e.g. sad, happy, neutral can be added to the application. This can also be implemented by using some machine leaning Algorithms.

REFERENCES

[1] Tapas Kumar Patra, Biplab Patra, Pushpangli Mohapatra, 2012 Text to Speech Conversion with Phonematic Concatenation, International Journal of Electronics Communication and Computer Technology.

[2] K.Partha Sarathy,AG Ramakrishnan, 2008 Text to speech synthesis system for mobile applications, 3rd language and technology conference

[3] Marc schroed, Jurgen Trouvain, 2003The German Te-to-Speech Synthesis System MARY: A Tool for Research, Development and teaching", Stuhlsatzenhausweg 3 D-66123 Saabr ucken Germany.

[4] M.Nageshwara Rao, Samuel Thomas, T.Nagarajan, and Hema A. Murthyarowsky, Text-to-Speech Synthesis using Syllable-like units, Indian Institute of technology,Madras.

[5] H.Hon,AAcerSIS UNITo,XHuang,J.Liu,and M.Plumpe, Automatic Generation Of Synthesis Units For Trainable Text-To-Speech Systems, Microsoft Research One Microsoft Way Redmond, Washington 98052,USA.

[6] Muhammad Masud Rashid,Md. Akter Hussain,M. Shahidur Rahman, Diphone Preparation for Bangla Text to Speech Synthesis.

[7] Margit Kastner,Brigitte Stangl, Exploring a text-to-speech feature by describing learning experience,enjoyment,learning styles,and values-Abasis for future studies",1530-1605/1226.00

[8] Sangam P. Borkar,Prof. S. P. Patil, "Text To Speech System For Konkani Language".

[9] Pradit Mittrapiyanuruk, Chatchawarn Hansakunbuntheung, Virongrong Tesprasit and Virach Somlertlamvanich,"Issues in Thai Text To Speech Synthesis: The NECTEC Approach "NECTEC technical