IMPLEMENTATION OF TEXT TO SPEECH WITH NATURAL VOICES FOR BLIND PEOPLE

Mr.D.BALAJI, M.E., Assistant Professor., MCE
Students Name:M.Rajukumar,P.Sethu,V.Hariharan,S.Sekar

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
MAHENDRA COLLEGE OF ENGINEERING
E-mail:balajiece21@gmail.com

Abstract: -**The objective of this paper is converting text into speech. A Text-to-speech synthesizer is an application that converts text into spoken word, by analyzing and processing the text using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. Here, we developed a useful text-to-speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3.file. The development of a text to speech synthesizer will be of great help to people with visual impairment and make making through large volume of text easier.**

Key words – text to speech, NLP, DSP, mp3.file, visual impairment

Introduction:

Today there is a wide spread talk about improvement of the human interface to the computer. Because no longer people want to sit and read data from the monitor. In this aspect Speech Synthesis is becoming one of the most important steps towards improving the human interface to the computer.In this world Disability of visual reading impacts the life of a person to a great extent. Text-to speech synthesizer (TTS) is the technology which lets computer speak to you. A blind person cannot also see the length of an input text when starting to listen it with the help of the speech synthesizer, an important feature is to give in advance some information of the text to be read successfully. Text-to-speech synthesis -TTS - is the automatic conversion of a text into speech that resembles, as closely as possible, a native speaker of the language reading that text. It is a computer-based system that should be able to read any text aloud.

At first sight, this task does not look too hard to perform. After all, is not the human being potentially able to correctly pronounce an unknown sentence, even from his childhood? We all have, mainly unconsciously, a deep knowledge of the reading rules of our mother tongue. They were transmitted to us, in a simplified form, at primary school, and we improved them year after year. However, it would be a bold claim indeed to say that it is only a short step before the computer is likely to equal the human being in that respect. Despite the present state of our knowledge and techniques and the progress recently accomplished in the fields of Signal Processing and Artificial Intelligence, we would have to express some reservations. As a matter of fact, the reading process draws from the furthest depths, often unthought-of of the human intelligence. The objective of this paper is to convert the English

text into speech. The conversion of English text into speech is done by using a stored speech signal data.

Proposed system:

Text to speech conversion module is designed by the use of mat lab. The recorded .wav (sounds) files are saved as a database separately. The phonemes are extracted from the text file. For text to speech conversion the concatenation method is proposed. The recorded speech are concatenated together to produce the synthesized speech. The resulting speech output is assessed by listening test.

BLOCK DIAGRAM:

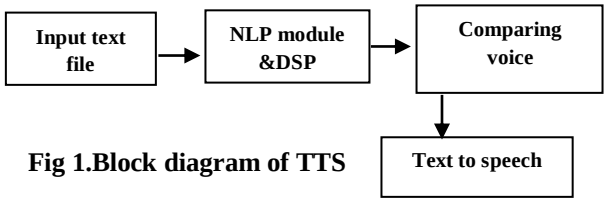


Fig 1.Block diagram of TTS

PROCESS:

Figure 2 introduces the functional diagram of a very general TTS synthesizer. As for human reading, it comprises a Natural Language Processing module (NLP), capable of producing a phonetic transcription of the text read, together with the desired intonation and rhythm (often termed as prosody), and a Digital Signal Processing module (DSP), which transforms the symbolic information it receives into speech. But the formalisms and algorithms applied often manage, thanks to a judicious use of mathematical and linguistic knowledge of developers, to short-circuit certain processing steps. This is occasionally achieved at the expense of some restrictions on the text to pronounce, or results in some reduction of the "emotional dynamics" of the synthetic voice (at least in comparison with human performances), but it generally allows to solve the problem in real time with limited memory requirements.

The NLP component:

Figure 2 introduces the skeleton of a general NLP module for TTS purposes. One immediately notices that, in addition with the expected letter-to-sound and prosody generation blocks, it comprises a morph-syntactic analyzer, underlying the need for some syntactic processing in a high quality Text-To-Speech system. Indeed, being able to reduce a given sentence into something like the sequence of its parts-of-speech, and to further describe it in the form of a syntax tree, which unveils its internal structure, is required for at least two reasons:

Accurate phonetic transcription can only be achieved provided the part of speech category of some words is available, as well as if the dependency relationship between successive words is known.

Natural prosody heavily relies on syntax. It also obviously has a lot to do with semantics and pragmatics, but since very few data is currently available on the generative aspects of this dependence, TTS systems merely concentrate on syntax. Yet few of them are actually provided with full disambiguation.

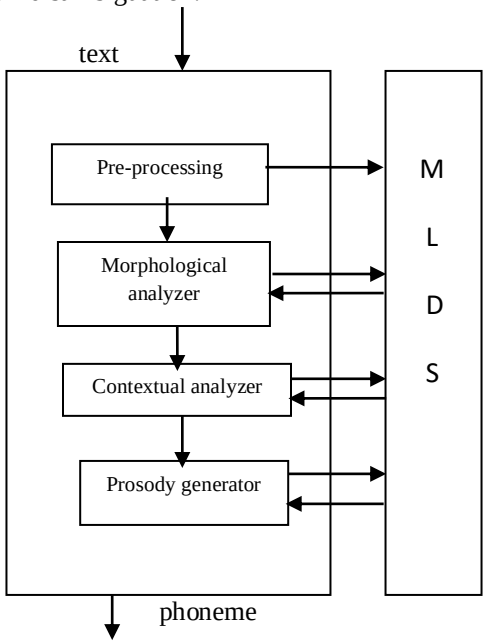


Fig 2.The NLP module

4.6.1Text analysis:

The text analysis block is itself composed of:

A pre-processing module, which organizes the input sentences into manageable lists of words. It

identifies numbers, abbreviations, acronyms and idiomatic and transforms them into full text when needed. An important problem is encountered as soon as the character level: that of punctuation ambiguity (including the critical case of sentence end detection). It can be solved, to some extent, with elementary regular grammars.

A morphological analysis module, the task of which is to propose all possible part of speech categories for each word taken individually, on the basis of their spelling. Inflected, derived, and compound words are decomposed into their elementary grapheme units (their morphs) by simple regular grammars exploiting lexicons of stems and affixes

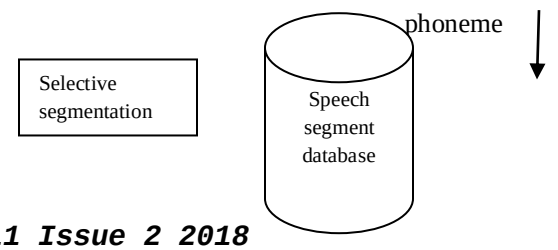
The contextual analysis module considers words in their context, which allows it to reduce the list of their possible part of speech categories to a very restricted number of highly probable hypotheses, given the corresponding possible parts of speech of neighboring words. This can be achieved either with n-grams, which describe local syntactic dependences in the form of probabilistic finite state automata (i.e. as a Markov model), to a lesser extent with mutli-layer perceptron's (i.e., neural networks) trained to uncover contextual rewrite rules, as in or with local, non-stochastic grammars provided by expert linguists or automatically inferred from a training data set with classification and regression tree (CART) techniques.

Finally, a syntactic-prosodic parser, which examines the remaining search space and finds the text structure (i.e. its organization into clause and phrase-like constituents) which more closely relates to its expected prosodic realization**.**

Database preparation:

A series of preliminary stages have to be fulfilled before the synthesizer can produce its first utterance. At first, segments are chosen so as to minimize future concatenation problems.

Digital signal processing module:



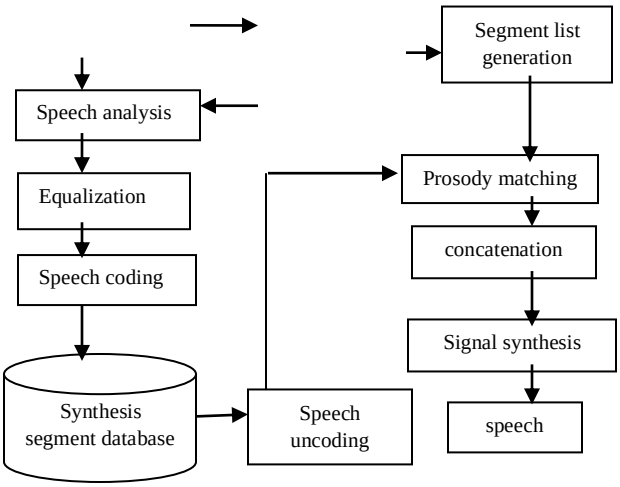


Fig 3.DSP Module

The operations involved in the DSP module are the computer analogue of dynamically controlling the articulator muscles and the vibratory frequency of the vocal folds so that the output signal matches the input requirements. In order to do it properly, the DSP module should obviously, in some way, take articulatory constraints into account, since it has been known for a long time that phonetic transitions are more important than stable states for the understanding of speech. This, in turn, can be basically achieved in two ways:

Explicitly, in the form of a series of rules which formally describe the influence of phonemes on one another;

Implicitly, by storing examples of phonetic transitions and co-articulations into a speech segment database, and using them just as they are, as ultimate acoustic units (i.e. in place of phonemes).

Two main classes of TTS systems have emerged from this alternative, which quickly turned into synthesis philosophies given the divergences they present in their means and objectives: synthesis-by-rule and synthesis-by-concatenation.

Challenges:

4.8.1 Text normalization challenges:

The process of normalizing text is rarely straightforward. Texts are full of heteronyms, numbers,and abbreviations that all

require expansion into a phonetic representation. There are many spellings in English which are pronounced differently based on context. For example, "My latest project is to learn how to better project my voice" contains two pronunciations of "project".

Text-to-phoneme challenges:

Speech synthesis systems use two basic approaches to determine the pronunciation of a word based on its spelling, a process which is often called text-to-phoneme or grapheme-to-phoneme conversion (phoneme is the term used by linguists to describe distinctive sounds in a language). The simplest approach to text-to-phoneme conversion is the dictionary-based approach, where a large dictionary containing all the words of a language and their correct pronunciations is stored by the program. Determining the correct pronunciation of each word is a matter of looking up each word in the dictionary and replacing the spelling with the pronunciation specified in the dictionary. The other approach is rule-based, in which pronunciation rules are applied to words to determine their pronunciations based on their spellings. This is similar to the "sounding out", or synthetic phonics, approach to learning reading.

Evaluation challenges:

The consistent evaluation of speech synthesis systems may be difficult because of a lack of universally agreed objective evaluation criteria. Different organizations often use different speech data. The quality of speech synthesis systems also depends on the quality of the production technique (which may involve analogue or digital recording) and on the facilities used to replay the speech. Evaluating speech synthesis systems has therefore often been compromised by differences between production techniques and replay facilities.

Prosodics and emotional content:

A study in the journal Speech Communication by Amy Drahota and colleagues at the University of Portsmouth, UK, reported that listeners to voice recordings could determine, at better than chance levels, whether or not the speaker was smiling. It was suggested that identification of the vocal features that signal emotional content may be used to help make synthesized speech sound more natural. One of the related issues is modification of the pitch contour of the sentence, depending upon whether it is an affirmative, interrogative or exclamatory sentence. One of the techniques for pitch modification uses discrete cosine transform in the source domain (linear prediction residual).

Advantage:

- Used for blind people
- It read the text fully like' HAPPY '
- It can read infinity words and sentences

Algorithm:

STEP 1: Create a database of various wave files.

STEP 2: Create a text file (.txt).

STEP 3: Open the .txt file in MATLAB.

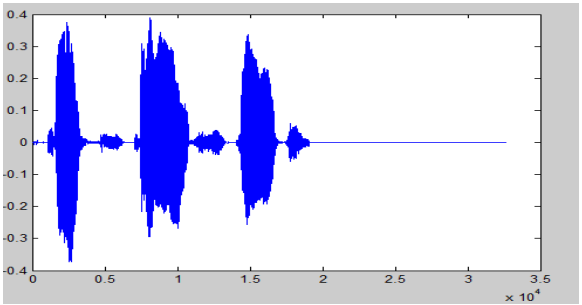
STEP 4: Read the file opened.

STEP 5: For each and every Character read and play corresponding wave (.wav) file.

Application:

- Aid to Vocally Handicapped
- Source of Learning for Visually Impaired
- Talking Books and Toys
- Games and Education.

Result:



As we have played the wave files corresponding to every character read, in character to voice conversion, we can also play the wave files for every word read. To record all the words of a

dictionary, the database memory also increased. Hence choosing sound unit with proper length isimportant, so that the word is natural and understandable when synthesized. Once the text is read, for every word the corresponding wave files are concatenated and played. The Text to Speech (TTS) conversion is performed using alphabets, numbers and words. For increasing the number of phonemes (.wav) the memory size also increases

Conclusion:

In this paper we discussed the development of TTS systems. This system is useful for blind people .In this work text into phonemes and then that phoneme into speech is converted by MATLAB. The limitation considers are, it does not read punctuation, romans number. Thus the system is very easy and efficient to implement and simple overview of working of text to speech system in step by step process. It gives a review about various concatenative speech synthesis method. The availability of standard databases speech synthesis frameworks and the comparison of different concatenative methods are discussed. The paper reviews the evaluation techniques for determining the quality of synthesized speech. In future work a miniaturized hardware implementation will be developed for helping visually impaired persons in understanding text they come across in day to day life.

REFERENCE:

1. R.S.S. Kumari, R. Sangeetha, Conversion of English Text To Speech Using Indian Speech Signal
2. N. Campbell, Evaluation of Speech Synthesis From Reading Machines to Talking Machines
3.M.Z. Rashad, H.M. El-Bakry, I.R. Isma'il, N. Mastorakis, An Overview of Text To Speech Synthesis Techniques
4.A. Trilla, Natural Language Processing techniques in Text-To-Speech synthesis and Automatic Speech Recognition, 2009
5.P.B. de Mareüil, C. d'Alessandro, A. Raake, G. Bally, M. Garcia, M. Morel A joint intelligibility evaluation of French text-to-speech synthesis systems: the EvaSy SUS/ACR campaign
6.E. Tzoukermann, Issues In Text To Speech For French
7.V. Delic, M. Sečujski, P.S.Molcer, Evolution of Text-to-Speech Systems and Methods of Their Assessment ji,..l,lti
8.P. Singh, A. Singh, A Text to Speech (TTS) System with English to Punjabi Conversion
9.S.P. Borkar, Prof. S.P. Patil, Text To Speech System for Konkani (Goan) Language
10.V. J. van Heuven, R. van Bezooijen, Quality Evaluation of Synthesized Speech