ORIGINAL RESEARCH PAPER

VOLUME-6 | ISSUE-5 | MAY - 2017 | ISSN No 2277 - 8179 | IF : 4.176 | IC Value : 78.46

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

CHARACTER RECOGNITION AND SPEECH SYNTHESIS SYSTEM USING LABVIEW

Engineering

G. Sowmiya

Assistant Professor, Department of Electronics and Communication and Engineering, Apollo Engineering College, Chennai.

ABSTRACT

The system has been implemented on Lab VIEW 7.1 platform. The developed system consists of OCR and speech synthesis. In OCR printed or written character documents have been scanned and image has been acquired by using IMAQ Vision for Lab VIEW. The different characters have been recognized using segmentation and correlation based methods developed in Lab VIEW

KEYWORDS:

Speech synthesis, dissecting speech, Character Recognition

1. INTRODUCTION

Knowledge extraction by just listening to sounds is a distinctive property and has become an important milestone in the evolution of species. Most of the animals are not only equipped with the means to extract information from the rich acoustical content of the environment and act accordingly, but they have the ability to produce sounds to interact with the environment as well. Humans have gone one step there, they have fairly advanced mechanisms that enable interaction within the species by very abstract rules of communication using voice – the language.

Recent progress in speech synthesis has produced synthesizers with very high intelligibility but the sound quality and naturalness still remain a major problem. However, the quality of present products has reached an adequate level for several applications, such as multimedia and telecommunications. With some audiovisual information or facial animation (talking head) it is possible to increase speech intelligibility considerably.

2. BACKGROUND

2.1 PHONETICS AND THEORY OF SPEECH PRODUCTION

Speech processing and language technology contains lots of special concepts and terminology. To understand how different speech synthesis and analysis methods work one must have some knowledge of speech production, articulatory phonetics, and some other related terminology.

Representation and analysis of speech signals

Continuous speech is a set of complicated audio signals which makes producing them artificially difficult. Speech signals are usually considered as voiced or unvoiced, but in some cases they are something between these two. Voiced sounds consist of fundamental frequency (F0) and its harmonic components produced by vocal cords (vocal folds). The vocal tract modifies this excitation signal causing formant (pole) and sometimes anti formant (zero) frequencies .Each formant frequency has also amplitude and bandwidth and it may be sometimes difficult to define some of these parameters correctly. The fundamental frequency and formant frequencies are probably the most important concepts in speech synthesis and also in speech processing in general.

With purely unvoiced sounds, there is no fundamental frequency in excitation signal and therefore no harmonic structure either and the excitation can be considered as white noise. Unvoiced sounds are also usually more silent and less steady than voiced ones. Whispering is the special case of speech. Speech signals of the three vowels (/a/ /i/ /u/) are presented in time- and frequency domain in Fig.2.1.

The fundamental frequency is about 100 Hz in all cases and the formant frequencies F1, F2, and F3 with vowel /a/ are approximately 600 Hz, 1000 Hz, and 2500 Hz respectively. With vowel /i/ the first three formants are 200 Hz, 2300 Hz, and 3000 Hz, and with /u/ 300 Hz, 600 Hz, and 2300 Hz. The harmonic structure of the excitation is also easy to perceive from frequency domain presentation.

It can be seen that the first three formants are inside the normal telephone channel (from 300 Hz to 3400 Hz) so the needed bandwidth for intelligible speech is not very wide. For higher quality, up to 10 kHz bandwidth may be used which leads to 20 kHz sampling frequency. Unless, the fundamental frequency is outside the telephone channel, the human hearing system is capable to reconstruct it from its harmonic components.

2.1.2 PROBLEMS IN SPEECH SYNTHESIS

The problem area in speech synthesis is very wide. There are several problems in text pre-processing, such as numerals, abbreviations, and acronyms. Correct prosody and pronunciation analysis from written text is also a major problem today.

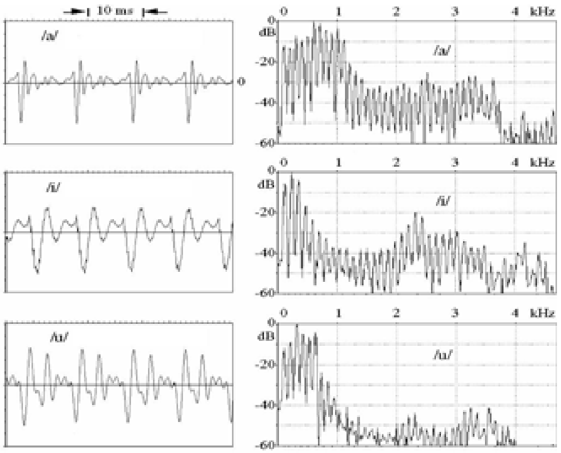


fig.2.1. The Time- And Frequency-Domain Presentation Of Vowels /a/, /i/, and /u/.

Written text contains no explicit emotions and pronunciation of proper and foreign names is sometimes very anomalous. At the low-level synthesis, the discontinuities and contextual effects in wave concatenation methods are the most problematic.

.3. LITERATURE SURVEY

When performing handwriting recognition on natural language text, the use of a word-level language model (LM) is known to significantly improve recognition accuracy. The most common type of language model, the n-gram model, decomposes sentences into short, overlapping chunks.In this paper, we propose a new type of language model which we use in addition to the standard n-gram LM. Our new model uses the likelihood score from a statistical machine translation system as a re ranking feature.[1]

Many text-to-speech synthesizers for Indian languages have used synthesis tech- niques that require prosodic models for good quality synthetic speech. However, due to unavailability of adequately large and properly annotated databases for Indian languages, prosodic models for these synthesizers have still not been developed properly.[2]

The recognized character information will be compared with the pre-defined data which we have stored in the system. As we are using the same font and size for the recognition there will be exact one unique match for the character.[3]

3.1 EXISTING SYSTEM

3.1.1 OPTICAL CHARACTER RECOGNITION

Optical character recognition (OCR) is the mechanical or electronic translation of images of hand-written or printed text into machine-editable text . The OCR based system consists of following process steps :

a) Image acquisition
b) Image pre –processing
c) Image segmentation
d) Matching and recognition

3. 2.1 IMAGE ACQUISITION

The image has been captured using a digital HP scanner. The image had been acquired using the program developed in LabVIEW as shown in the Fig.3.1. The configuration of the Image has been done with the help of Imaq create subvi function of LabVIEW. The configuration of the image means selecting the image type and border size of the image as per the requirement. In this work 8 bit image with border size of 3 has been used.

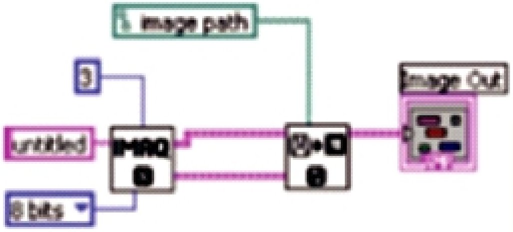


Fig.3.1. Image Acquisition.

3.2.2 Image Pre-processing (Binarization)

Binarization is the process of converting a gray scale image (0 to 255 pixel values) into binary image (0 to1 pixel values) by using a threshold value. In this work, a global thresholding with a threshold value of 175 has been used to binarize the image i.e. the values of pixel which are from 175 to 255 has been converted to 1 while the of pixel which have gray scale value less than 175 have been converted to 0. The LabVIEW program of binarization has been shown in Fig.3.2

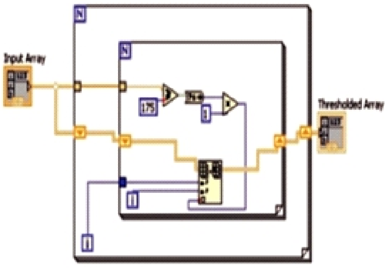


Fig.3.2. Binarization Process.

3.2.3. Word Segmentation

In the word segmentation process the line segmented images have been vertically scanned to find first ON pixel. When this happen the system remember the coordinate of this point as x1. This is the starting coordinate for the word.. The system records the first OFF pixel as x2. From x1 to x2 is the word. The Figure 3.4 shows the Lab VIEW programs of word segmentation

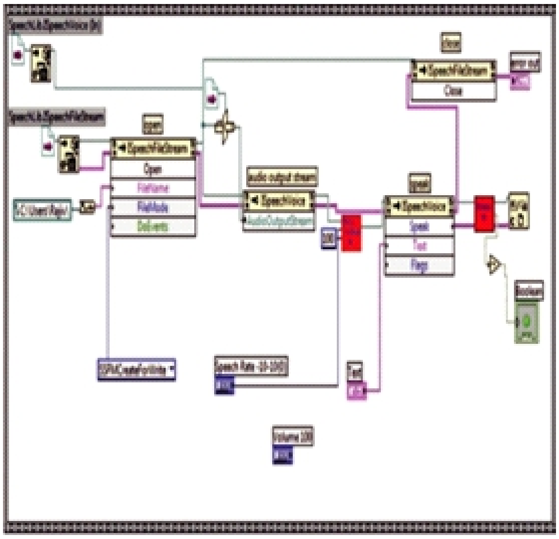


Fig.3.3 Word Segmentation

4. PROPOSED SYSTEM

4..1 TEXT TO SPEECH SYNTHESIS

In text to speech module text recognized by OCR system will be the inputs of speech synthesis system which is to be converted into speech in .wav file format and creates a wave file named output wav, which can be listen by using wave file player.

4.2 TEXT TO SPEECH CONVERSION

In the text speech conversion input text is converted speech (in Lab VIEW) by using automation open, invoke node and property node. Lab VIEW program of Text to speech conversion is shown in Figure

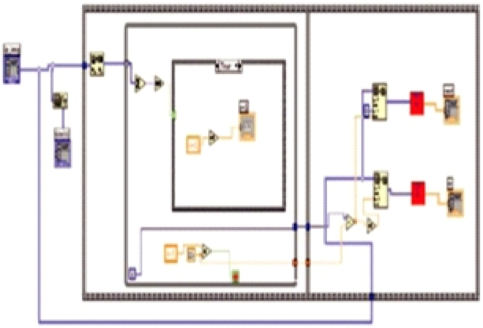


Fig.4.1 Text to Speech Conversion

In Text to Speech system the hardware requirements are very less. It requires only a good quality speaker for the production of sound signal. The software part is developed using the LabVIEW Software.

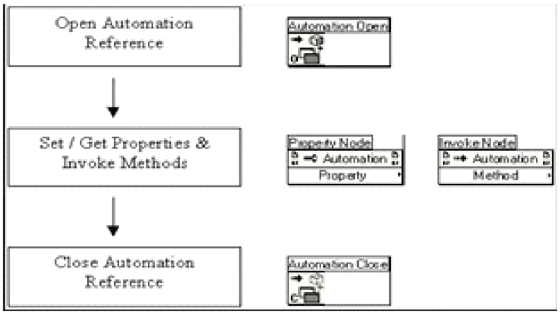
Here, in text to speech module a text box is created so that user can write text which is to be converted into speech in .wav file format and creates a wave file named output .wav, which can be listen by using wave file player.

4.3 Activex and Lab VIEW

ActiveX is the general name for a set of Microsoft Technologies that allows users to re-use code and link individual programs together to suit their computing needs. Based on COM (Component Object Model) technologies, ActiveX is an extension of a previous technology called OLE (Object Linking and Embedding).

4.4 Lab VIEW As An Automation Client

Lab VIEW provides functions in its Application Programmable Interface (API) that allow it to act as an automation client with any automation server. The figure 4.1 shows the programming flow used in Lab VIEW, and gives the associated functions with each block.



In general, information about a program's ActiveX automation server can be obtained from the program's documentation or by browsing the program's type library.

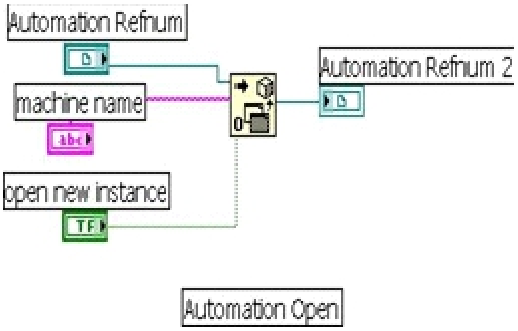


Fig.4.1 Programming flow of ActiveX used in Lab VIEW

4.4.1 Automation Open

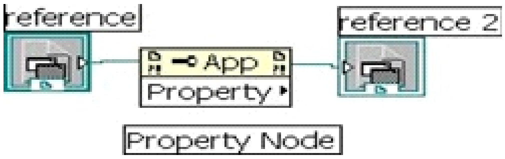
Returns an automation refnum, which points to a specific ActiveX object. In Text to Speech VI, it gives refnum for Microsoft speech object library

4.5 Invoke Node

Invokes a method or action on a reference. Most methods have associated parameters. If the node is configured for VI Server Application class or Virtual

4.6 Property Node

Gets (reads) and/or sets (writes) properties of a reference. The Property Node automatically adapts to the class of the object that you reference.



5. CONCLUSIONS

The LabVIEW tool allows us to implement the Text to Speech conversion .LabVIEW has its strong inbuilt Speech library to implement the Text to Speech conversion.

Various methods & techniques are available for speech synthesis which is discussed in this report. Here, in text to speech conversion VI ,a text box is created so that user can write text which is to be converted into speech in .wav file format and creates a wave file named output .wav , which can be listen by using wave file player.

5.1 FUTURE SCOPE

A speech synthesize system i.e. a text to speech conversion system is developed using the LabVIEW. Still some more work can be done in this field as mentioned below:

1) By adding some reverberation it may be possible to increase the pleasantness of synthetic speech afterwards
2) Different sound wave format files such as .mp3 etc. or other format required by user, can produced using different techniques
3) Different methods for correct pronunciation can be used to improve better and correct pronunciation.
4) Provision for controlling the bits/samples, and hence the speech speed, pitch control etc can be added as new feature.

A new module can be added for the voice activated remote control applications

REFERENCES

1. Sproat, Richard W., and Joseph P. Olive. "A modular architecture for ultilingual text-to-speech." In Progress in speech synthesis, pp. 565-573. Springer New York, 1997.
2. Beskow, Jonas, KjellElenius, and Scott Mc Glashan. "The OLGA project: An animated talking agent in a dialogue system." In Proceedings of Eurospeech, vol. 97. 1997.
3. Davaatsagaan, Munkhtuya, and Kuldip K. Paliwal. "Diphone- Based Concatenative Speech Synthesis System for Mongolian." In Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1. 2008.
4. Cowie, Roddy, and Ellen Douglas-Cowie. "Automatic statistical analysis of the signal and prosodic signs of emotion in speech." In Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, vol. 3, pp. 1989-1992. IEEE, 1996.
5. Chasin, Marshall. "Published on Thursday, 01 September 2011 07:59."
6. Flanagan, James Loton. Speech synthesis. Vol. 3. Dowden Hutchinson and Ross, 1973.
7. O'shaughnessy, Douglas. Speech communication: human and machine. Universities press, 1987.
8. Breen, A. P., E. Bowers, and W. Welsh. "An investigation into the generation of mouth shapes for a talking head." In Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, vol. 4, pp. 2159-2162. IEEE, 1996.
9. Veldhuis, Raymond NJ, I. J. M. Bogaert, and N. J. C. Lous. "Two-mass models for speech synthesis." In Fourth European Conference on Speech Communication and Technology. 1995.