

READER: DEVICE TO CONVERT WRITTEN/PRINTED TEXT TO SPEECH

Shahbaaz Ahamad S.A¹, Harshitha P.K²

¹Department of Electronics & Communication, NMAMIT, Nitte, Karkala, Karnataka, India

²Department of Electronics & Communication, NMAMIT, Nitte, Karkala, Karnataka, India

Abstract

A device which is developed with the purpose of helping the blind and the old aged to be able to read books, articles, magazines or news papers. Using the ICR and TTS technology the written format is to be converted into speech which the user listens to with the help of earphone connected to 3.5 mm jack.

Keywords: Scanner, ICR, OCR, SPRAM, SDRAM, TTS

1. INTRODUCTION

Blind people are incapable of reading books which do not have Braille. This problem can be disentangled if the printed matter by any means is converted to a form which can be inferred by the blind, the finest would be speech.

The above mentioned solution can be applied if the device used for the processing of the text to a voice signal, is a portable one i.e. the user (blind person) must be able to carry it to any place just like he carries his mobile phone. The first part would be a scanner. There are hand held scanner's which have already made their mark in the field of scanning a document or a photograph with the accuracy comparable to the desktop scanner. The scanners which use the optical character recognition system (OCR) powered processor can be used for the detection of text in the document scanned. The entire processing requires a SOC (system on chip). In a SOC most of the components of a device are placed in an extremely small area on the device, as a result the size of the board bearing the components reduces, furthermore it helps to reduce costs for assembling the product compared to many parallel multi-chip setups. An embedded SOC integrates the IP cores for processors, memories, and application-specific hardware to create the SOC. Here SOC includes an intelligent character recognition (ICR) based scanning support, scratch pad random access memory SPRAM, synchronous dynamic random access memory SDRAM and text to speech TTS system embedded on it.

The physical size of this design can be expected to be portable since the major part included 'the scanner' is hand held. Though the major aim is to eradicate the problem of reading books for blind the device can be further used by even the old aged to read the text they want to easily by not reading it but hearing it.

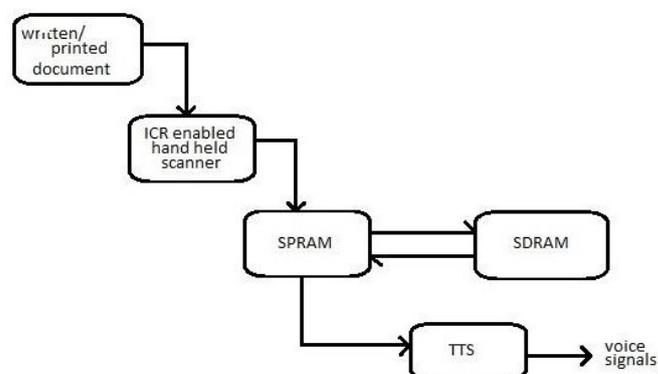


Fig.1. various stages of the READER

As mentioned earlier the data is scanned using the scanner and then transferred to SPRAM where it is further processed. Then the data is moved to TTS system where it gets converted to a speech signal.

2. ICR ENABLED HAND HELD SCANNER

The basic aim of using a scanner is to capture the image of the document. A hand held scanner uses a fluorescent lamp in the scan head (we can even use cold cathode fluorescent lamp or L.E.D lamps). The scanner will have a charge coupled device (CCD) which consists of light sensitive diodes that convert light energy to electrical energy. So now how the scanner works! There is a couple of mirrors placed in the scan head facing the document when the fluorescent lamp emits light, the reflected image falls on the mirrors placed. The mirror in the which is the final one to sense the signals, sends them to the lens. The lens passes on the light rays into a filter which is present in charge coupled device (CCD) and then for further processing it is sent to the processor where further processing is done depending on the electrical signals received. There are series of processing steps involved and all of which is present on the circuit board which will be used.

Optical character recognition (OCR) is a system to obtain the electronic identification of printed alphanumeric characters by scanning. Printed text comprising characters is scanned

and stored as an image, further the characters are recognized by the OCR engine and then the printed characters are converted to machine readable form (ASCII). OCR enables the users to swiftly obtain the digitized form of the printed text that lets the user to alterate the text electronically. OCR is of two types first Optical character recognition - recognizes one character at a time, second Optical word recognition - recognizes one entire word at a time. The development and research in the field of OCR led to another recognition system termed ICR(Intelligent Character Recognition).

Intelligent Character Recognition (ICR) has the capability to extract electronic ASCII data from printed text more than that ICR can also take out the ASCII equivalent from handwritten materials. The major difference between an OCR and ICR system is that ICR system can recognize the handwritten matter better than the OCR system. The registration marks on the four corners of a document are used by the ICR/OCR to recognize an image on a page. OCR/ICR is flexible as timing tracks or blocks (like in ID's) is not required. Scanners output can be elevated by the use of drop color.

2.1 OCR/ICR Processing Stages

Scanning Process

Document Scanning is the first step involved in the character recognition both by ICR as well as OCR. Time taken for scanning will be dictated by the nature of the scanner machines, shading, paper quality, cleanliness, weights, right setting of the ICR/OCR framework.

Recognition Process

The perceiving procedure is to translate pictures, in other words identify and investigate images. The right memory and the setup limit will decide the precision of elucidation of the ICR/OCR

Verification Process

To analyze the estimation of the translated picture with the genuine picture of the structure

3. MEMORY SYSTEM

Innovation has created up to a degree of coordinating a few processor center on the same chip with installed recollections and circuits. The best case in the single chip of cellular telephone it incorporates modem, radio handset, power administration usefulness, a mixed media motor ,security highlights and numerous all the more all on the same chip.Thus here we make use of a similar kind of embedded system which is application specific i.e. to read characters or words and convert it to speech .an embedded system has a memory on chip which is SPRAM(scratch pad RAM)or sometimes referred to as cache. Purely cache-based on-chip memory organization is generally not preferred for an embedded system as that organization cannot guarantee the worst-case execution time. Hence an external memory must be used for storing of data so we use a SDRAM

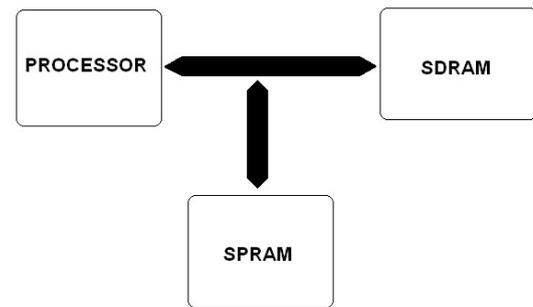


Fig.2. memory management

4. SDRAM

SDRAM, stands for Synchronous dynamic random access memory, in the following context we use the SDRAM to store the data (characters and voice) which will used later based on the necessity. RAM comes in sticks, or modules, which comprise of a few little chips that store information. on the off chance that you have additional RAM spaces, you can add more memory to your framework without evacuating the current memory, up to the greatest sum that your installed framework bolsters. there are an assortment of them accessible.

- One of the earlier RAM protocols was SDRAM, and it's almost obsolete .Typically, compared to hundreds of options in other types of RAM. SDRAM is 168-pin RAM, and runs at speeds ranging from 66MHz to 133MHz in the following context we use the SDRAM to store the data (characters and voice) which will used later based on the necessity.
- **DDR SDRAM is the modern starting point:** DDR SDRAM is short for Double Data Rate Synchronous DRAM and is like however speedier than the old SDRAM standard. DDR RAM is a 184-pin module, and works fine. It's an older RAM protocol and it'll go the way of SDRAM eventually. You may still have as many as a hundred options for DDR RAM, whereas you'd have over 300 options for newer memory protocols. In technical speak; DDR RAM completes two data transfers per memory clock cycle.
- **DDR2 RAM is a more common modern RAM selection:** DDR2 RAM is the modern computer memory protocol, and it is the typical choice for a general workstation. It's a 240-pin RAM module, and you guessed it-it's faster than DDR RAM. Through innovations in technology, DDR2 RAM completes four data transfers per memory clock cycle, making it considerably faster than DDR RAM.
- **DDR3 RAM is the next generation of RAM:** DDR3 RAM is the newest memory protocol and is effectively twice as fast as DDR2 RAM .While DDR3 RAM is a 240-pin module; it's not backwards compatible with DDR2 or any prior technology.

5. SCRATCH PAD RAM (SPRAM)

SPRAM is a high speed internal memory used for temporary storage of data which needs to be processed to obtain the required results. The real-time constraints of embedded applications and its need to run the applications in real-time

requires on-chip SPRAM. Scratch pad refers to a memory circuit used to hold small amount of data for a rapid access using a high speed microprocessor. A system with scratch pad can be assumed to be a system with non-uniform memory access latencies. SPRAM based on-chip memory organization is suitable only for less complex embedded applications. It is possible to allocate different data sections in SPRAM at different points by moving data dynamically between the off-chip memory and SPRAM. A proper placement of the data sections on the SPRAM is required to ensure that the SPRAM misses are reduced and throughput is increased effectively. The on-chip SPRAM architecture constitutes the memory organization and has a significant effect on the area, performance and power consumption. The data between on-chip SPRAM and external RAM can be partitioned when the space available in 'on-chip SPRAM' is less than required for processing of a particular word or information. Data partitioning aims at improving the overall performance of the memory sub-system by placing data in a particular manner in the memory so that we get:

- Higher access frequency
- Data that has an overlapping life time with many other data
- Data that has poor spatial access characteristics

Data that exhibits the above characteristics can be placed in SPRAM which reduces the cache misses and also the coherence problem thus leading to an overall improvement in the memory sub-system performance. Unlike DSP processor's on-chip SPRAM, the microcontroller's on-chip RAM is not organized as multiple memory banks. This is because the MCU applications do not require intensive computations.

6. TEXT TO SPEECH(TTS)

Normal language text is converted into speech by means of a TTS system. The design of text-to-speech systems are distributed as front end and back end. In the front end, various text symbols and characters are converted into a list of equivalent words. This method is referred to as text-normalization, pre-processing or tokenization. After pre-processing, the back end then converts these words into sound. The text information is synthesized by a digital signal processing module into speech. This process is called speech synthesis. TTS systems are provided by third party vendors which are made available in different languages and vocabularies depending on the user specifications. Text and linguistic analysis are performed during the front end and back end design of the TTS system.

Text Analysis: The data stream is broken into words, phrases, symbols, or other meaningful elements called tokens. The list of tokens becomes input for further processing.

Linguistic Analysis: It is a method to analyze a given language sample. This approach is used to understand and describe the ability to speak a given language. It is a technique to interpret how the human mind processes and creates language. Phrasing, intonation and duration are the three terms associated with a speech pattern in Linguistic analysis. Phrasing is dividing speech into segments by modifying the speech. Intonation is the style of utterance of words which is basically the variation in the pitch of voice. Duration refers to the duration pattern of speech. this pattern is different for different languages and different styles of a particular language. This results in a complex duration model.

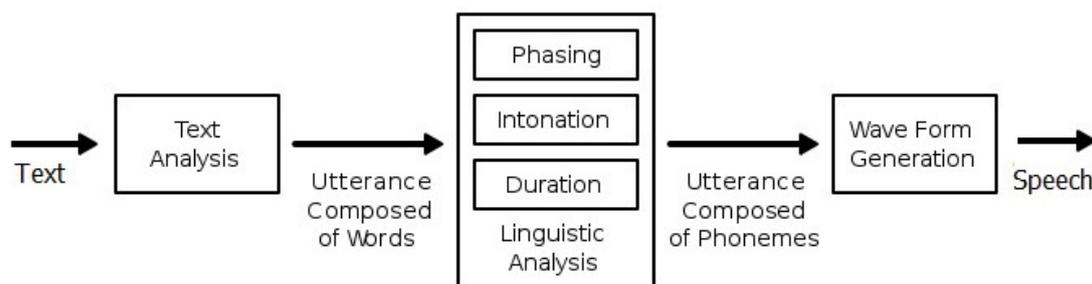


Fig 3. Overview of a typical TTS system

Speech synthesis is a part of linguistic analysis and an output from a computer or any device obtained after processing the text input. The output is basically reading out the words which involves three stages of processing. The first stage in speech synthesis is normalization of text to words. For example, the number 501 can be read as five hundred and one when it refers to a quantity or may be read as five four one in a case of a padlock combination where each number is read differently. The next stage is conversion of word to phonemes. Phoneme is a form of sound that distinguishes two words. There are 44 phonemes in the UK English model each of which represents a different sound. The last stage is to compile these phonemes to sound. At this

stage, the sound to be produced is pre-loaded into the computer as human recorded voices.

7. FUTURE SCOPE

An attempt can be made to install this feature in the existing hand held scanners, thus reducing the extra hardware requirements. ICR can be updated to Intelligent word recognition(IWR) which can recognize not only printed handwritten information, but cursive handwriting as well. ICR recognizes on the character-level, whereas IWR works with full words or phrases. IWR is said to be more evolved than hand print ICR. IWR is an alternate to ICR and OCR

engines. It is optimal for real-world processing of documents such as hand written text which is hard to recognize for such engines. With IWR, error associated with character recognition is reduced.

8. CONCLUSION

In this paper we have described READER, a portable gadget which enables the blind to supersede the use of Braille. It is an ingenious way to help the blind read non-Braille books. READER uses an Application specific SOC and the application involved here is only to read and convert to speech. The application program in modern embedded system is complex since most of them support a variety of device interfaces such as network interfaces, credit card readers, USB interfaces, parallel ports and other similar interfaces. As the embedded system used in this device is less complex with few interfaces the gadget performance is winged!

The ICR technology used in READER improves the accuracy of the device as hand written material can also be read (however, the threshold value is considered). Unlike OCR, ICR gives the speech synthesis modules the ability to compile hand written text. The ability to capture cursive makes this method unique.

In this paper we introduced READER, presented information about its scanner enabled with ICR, memory organization and speech synthesis. We believe that READER if designed will be a remarkable innovation and meet our purpose to help the blind.

ACKNOWLEDGEMENT

We would like to express our gratitude to Dr. Rekha Bhandarkar, Head of Department, Electronics and Communication Engineering, N.M.A.M. Institute of Technology for her reviews, motivation and the valuable suggestions, which significantly improved this article.

REFERENCES

- [1]. Gerald F Marshall, handbook of optical and laser scanning
- [2]. Preeti Ranjan Panda, B. V. N. Silpa, Aviral Shrivastava, Krishnaiah Gummidipudi, Power-efficient System Design
- [3]. T.S. Rajesh Kumar, On-Chip Memory Architecture Exploration of Embedded System on Chip
- [4]. Thierry Dutoit, An Introduction to Text-to-Speech Synthesis