Developing a Portable Reading Machine for the Blinds

Erwin Normanyo [1,*,] Dzinyefa Kodzo Tsolenyanu [2], Isaac Adetunde [3]

1.   University of Mines and Technology, Department of Electrical and Electronic Engineering, Tarkwa, Ghana
2.   Process and Plant Automation, P. O. Box SR 95, Accra, Ghana
3.   University of Mines and Technology, Department of Mathematics, Tarkwa, Ghana
E-mail address: enormanyo@umat.edu.gh

Abstract: Prevalence of the causes of blindness, especially cataract, is alarming and most sighted people may get it in their old age. Most blind do not have access to readily transliterated documents such as instructions on food packaging, medication and newspapers. Hitherto existing touch-based methods of Moon and Braille for text cognition by the blind and visually impaired are no longer acceptable technologically. In this work, a reading machine for the blind and visually impaired has been developed enabling them to read novels, newspapers, books and letters. In the development, scanner, optical character recognition, and text-to-speech technologies were employed. The Fourier Transform was involved in signal and image processing. Software implementation made use of XML-based speech synthesis markup language. Orientation of the document/paper does not matter during the scanning process. The SSML (Natural reader software) can still identify the right position of words and read them in a natural sounding voice. Li-Ion batteries used give high energy density and higher voltage ensuring reliability. With the implementation of the reading machine developed, information should be carried indiscriminately to the blind and visually impaired [The Journal of American Science. 2010;6(1):1-14]. (ISSN: 1545-1003).

Key words: the blind; portable reading machine; natural reader software; Fourier transformation; visually impaired.

1.   Introduction

Blindness is particularly devastating in the developing world where it has a profound impact on the quality of life for the blind person and his or her community. Life expectancy of the blind is usually less than half that of someone with eyesight the same age. The desperateness of this situation is augmented by the fact that a blind person is unable to contribute to the family income. Not only does blindness mean a father is unable to work, or a mother cannot collect water or go to market, but someone with eyesight must care for him or her. Effectively two income producing individuals are lost. This creates a devastating economic impact on the family and the community. Restored eyesight allows the individual to return to a normal life of work and a traditional role in the family.

In Ghana, about 4.4% of the population is blind and people above the age of 50 years experience low vision. Pitifully enough, many novels, newspapers, books and letters are not readily transliterated into Braille to convey the information to the blind. Means of communication between the sighted and the blind is chiefly vocal. Therefore, the need for a reading machine is paramount. Out of the 20 million people living in Ghana, it is estimated that 200,000 are blind and over 600,000 more people are visually impaired. Thus, blindness is affecting about 4.4 % of the Ghanaian population and people beyond the age of 50 years experience low vision (Dogbe, 2004). A cross-sectional drawing of the eye is given in Figure 1.

Blindness is the total or partial inability to see due to disease or disorder of the eye, optic nerve, or brain (Microsoft Encarta, 2007). The term blindness typically refers to vision loss that is not correctable with eyeglasses or contact lenses (Microsoft Encarta, 2007). Blindness may not mean a total absence of sight, because, some people who are considered blind may be able to perceive slowly moving lights or colors. The term low vision is used for moderately impaired vision. People with low vision may have a visual impairment that affects only central vision (the area directly in front of the eyes) or peripheral vision (the area to either side of and slightly behind the eyes). Some people with low vision are able to function with their remaining sight while others need help to learn to use their sight more efficiently with training and special tools

Color blindness, for example, does not reduce visual acuity and should more accurately be called color-perception deficiency. Color blindness occurs almost exclusively in males, and the most common form is the inability to differentiate between certain shades of red and green. Night blindness, the inability to see in low levels of light, is commonly associated with a lack of vitamin A in the diet or with inherited diseases such as retinitis pigmentosa, a condition involving progressive degeneration of the eye's retina and abnormal deposits of pigment. In Ghana about a million people are blind (Dogbe, 2004).
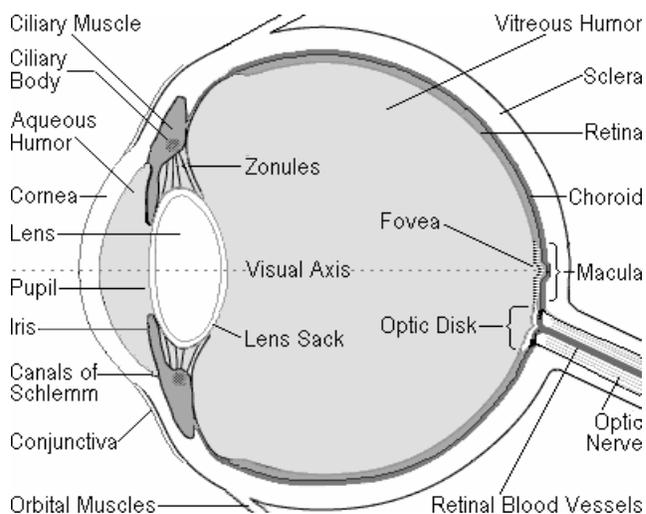
Figure 1. Cross sectional drawing of the eye (side view)

In spite of the progress made in surgical techniques in many countries during the last ten years, cataract (47.9 %) remains the leading cause of visual impairment in all areas of the world, except for developed countries (WHO, 2009a). With the exception of age-related macular degeneration (AMD), the rest are the causes of avoidable visual impairment worldwide. However, in developed countries, AMD is the leading cause of blindness, due to the high life expectancy of over 70 years of age. In the least-developed countries, and in particular Sub-Saharan Africa, the causes of avoidable blindness are primarily, cataract (50 %), glaucoma (15 %), corneal opacities (10 %), trachoma (6.8 %), childhood blindness (5.3 %) and onchocerciasis (4 %) (WHO, 2009a). In Table 1 is given the global estimate of visual impairment.

Figure 2. An eye with cataract (Source: Microsoft Encarta reference library, 2007)

1.1 Causes of Blindness and Visual Impairment
1.1.1   Cataracts of the Eye
Cataracts are formed in the lens of the eye which is behind the black dot (pupil) in the middle of the eye. It is a clouding of the lens, which prevents a clear, sharp image being produced. A cataract forms because the lens is sealed in a capsule (pupil as shown in Fig. 1) and as old cells die they get trapped in the capsule, with time this causes a clouding over of the lens (Fig. 2.). This clouding results in blurred images. This is when the lenses become opaque meaning that no light goes through.

1.1.2   Glaucoma of the Eye
Another disease is called glaucoma (Figure 3). The most common type of this disease occurs in people who are 40 years or older and the other type occurs in babies when they are born. The eye produces a clear fluid (aqueous humor) from the lacrimal gland that fills the space between the cornea and the iris as shown in 3. This fluid produces tears to clean, moisten and lubricate the eyes and then drains the excess fluid into the nose through a complex drainage system. It is the balance between the production and drainage of this fluid that determines the eyes intraocular pressure (IOP).
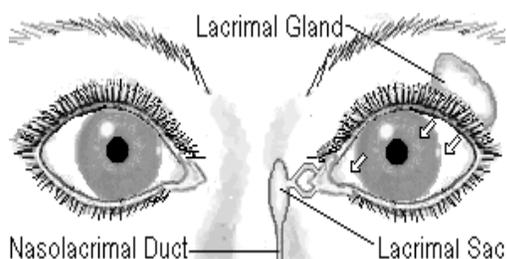
Figure 3. Glaucoma of the eye

1.1.3 Trachoma of the Eye
Trachoma popularly known in Ghana as "Apollo" is one of the oldest infectious diseases known to mankind. It is caused by Chlamydia trachomatis – a micro organism which spreads through contact with eye discharge from the infected person (on towels, handkerchiefs, fingers, etc.) and through transmission by eye-seeking flies. After years of repeated infection, the inside of the eyelid may be scarred so severely that the eyelid turns inward and the lashes rub on the eyeball, scarring the cornea (the front of the eye). If untreated, this condition leads to the formation of irreversible corneal opacities and blindness.

1.1.4 Age-Related Macular Degeneration
Macular degeneration makes people not see things at the center of their field of vision. This is a degenerative condition of the macula (the central retina). It is caused by the hardening of the arteries that nourish the retina. This deprives the retinal tissue of the nutrients and oxygen that it needs to function and causes deterioration in central vision. This disease cuts off the circulation of blood in the center of the retina. It can be treated with a laser. This loss of sight often occurs as people's age increases.

1.1.5 Diabetic Retinopathy
Diabetic retinopathy happens to people who have diabetes mellitus for a few years. Diabetes changes the blood vessel of the retina. The retina is the part of the eye that absorbs light rays. Sometimes the blood vessels will burst and cause bleeding in the eye. Sometimes the retina is detached from the back of the eye. Another case is when fluid leaks from capillaries in the retina. If your retina is detached or you have bleeding in the eye the clear fluid fills the center of the eye that can cause blindness.

1.2 Distribution of Visual Impairment
Visual impairment distribution is done according to age, gender, and geographical location factors.

1.2.1 Age
Visual impairment is unequally distributed across age groups. More than 82 % of all people who are blind

are 50 years of age and older, although they represent only 19% of the world's population. Due to the expected number of years lived in blindness (blind years), childhood blindness remains a significant problem, with an estimated 1.4 million blind children below age 15 (WHO, 2009b).

1.2.2 Gender
Available studies consistently indicate that in every region of the world, and at all ages, females have a significantly higher risk of being visually impaired than males (WHO, 2009b).

1.2.3 Geographical Location
Visual impairment is not distributed uniformly throughout the world. More than 90% of the worlds visually impaired live in developing countries (WHO, 2009b).

1.3 Reading Techniques
Reading is an activity characterized by the translation of symbols, or letters, into words and sentences that have meaning to the individual. The ultimate goal of reading is to be able to understand written material, to evaluate it, and to use it for one's needs. Reading exposes people to the accumulated wisdom of human civilization. Mature readers bring to the text their experiences, abilities, and interests; the text, in turn, allows them to expand those experiences and abilities and to find new interests. In order to read, one must follow a sequence of characters arranged in a particular spatial order. For example, English flows from left to right, Hebrew from right to left, and Chinese from top to bottom. The reader must know the pattern and use it consistently.

Ordinarily, the reader sees the symbols on a page, transmit the image from the eye to the brain and pronounce them in the mind or aloud through the vocal cavity. However, reading techniques for the blind namely the Moon and the Braille are quite different from the sighted person. The technique employed by a blind is shown in Figure 4.
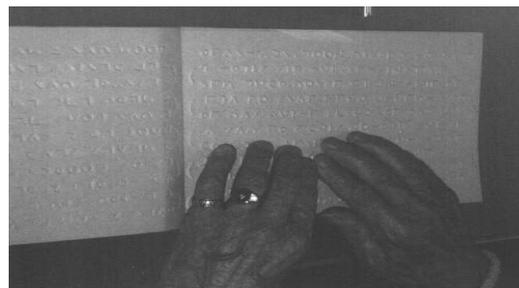

Figure 4. A person reading moon or braille.
(Source: Microsoft Encarta library, 2007)

Table 1. Global estimate of visual impairment

	African Region	Region of the Americas	Eastern Mediterranean Region	European Region	South-East Asia Region	Western Pacific Region	Total
Population	672.2	852.6	502.8	877.9	1,590.80	1,717.50	6,213.90
Number of blind people	6.8	2.4	4	2.7	11.6	9.3	36.9
Percentage of total blind	18 %	7 %	11 %	7 %	32 %	25 %	100 %
Number with low vision	20	13.1	12.4	12.8	33.5	32.5	124.3
Number with visual impairment	26.8	15.5	16.5	15.5	45.1	41.8	161.2

(Source: WHO, 2009b)

1.3.1 Braille

Braille is a writing system which enables blind and partially sighted people to read and write through touch. It was revised by Louis Braille (1809-1852), a French teacher of the blind. It consists of patterns of raised dots arranged in cells of up to six (6) dots in a 3 x 2 configuration as shown in Figure 5. Braille has been adapted to writing many different languages including even Chinese, and is also used for musical and mathematical notations. Each cell represents a letter, numeral or punctuation mark. Some frequently used words and letter combinations also have their own single cell patterns.
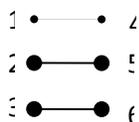

Figure 5. Six dots in 3 × 2 configuration

Braille can be categorized into the grades 1, 2, and 3. Grade 1 consists of the 26 standard letters of the alphabet and punctuation. It is only used by people who are first starting to read Braille. Secondly, grade 2 consists of the 26 standard letters of the alphabet, punctuation and contractions.

The contractions are employed to save space because a Braille page cannot fit as much text as a standard printed page. Books, signs in public places, menus, and most other Braille materials are written in Grade 2 Braille. Last but not least grade 3 which is used only in personal letters, diaries, and notes. It is a kind of shorthand, with entire words shortened to a few letters.

(1) Formation of Letters of the Alphabet in Brail

The formation of letters of the alphabet is best organized as: letters from A – J which are the first ten (10) upper dots followed by the letters from K – T which are letters formed by adding dot three (3) to each of the first ten letters, letters of from U – Z are formed by adding dot six (6). Table 2 gives the summary of table representation of basic letters and abbreviations of some words. Braille representation of words and abbreviations is presented in Table 3.

Table 2. Summarized table representation of basic letters

•	:	••	•:	•.	••	••	:.	.•	.•	•	:	••
										•	•	•
a	b	c	d	e	f	g	h	i	j	k	l	m
:•	•.	:•	:•	:•	.•	.•	•	:	.•	••	••	•.
•	•.	•	•	••	•	••	••	:.	••	••	••	••

(Source: Anon, 1999a)

Table 3. Braille representation of words and abbreviations

•	:	••	•:	•.	••	••	:.	.•	•	:	••	••
									•	•	•	•.
a	but	can	do	every	from	go	have	just	knowledge	like	more	not
••	••	:•	.•	.•	•	:.	.•	••	••	•.	••	••
•	••	•.	•	•.	••	:.	••	••	••	••	••	••
	•											
people	quite	rather	so	that	us	very	will	it	you	as	and	for
•.	.•	.•	•	•	••	••	•.	••	••	•.	.•	•
••	•.	••	•.	•.	•.	:•	•.	•.	••	••	•.	•
												•
of	the	with	child/ch	gh	shall/sh	this/th	which/wh	ed	er	out/ou	ow	bb
••	••	•.	••	••	•	.•	••					
	••	•.	••	.•	•	•.	.•					
						••	•					
cc	dd	en	gg; were	in	st	ing	ar					

(Source: Anon, 1999a)

(2) Sample Texts in Braille
The Braille text below in Figure 6 is transliterated to mean, "Be kind to others"

Figure 6. Braille representation of "Be Kind to Others"

Braille text in Figure 7 below is the article 1 of the universal declaration of human rights.
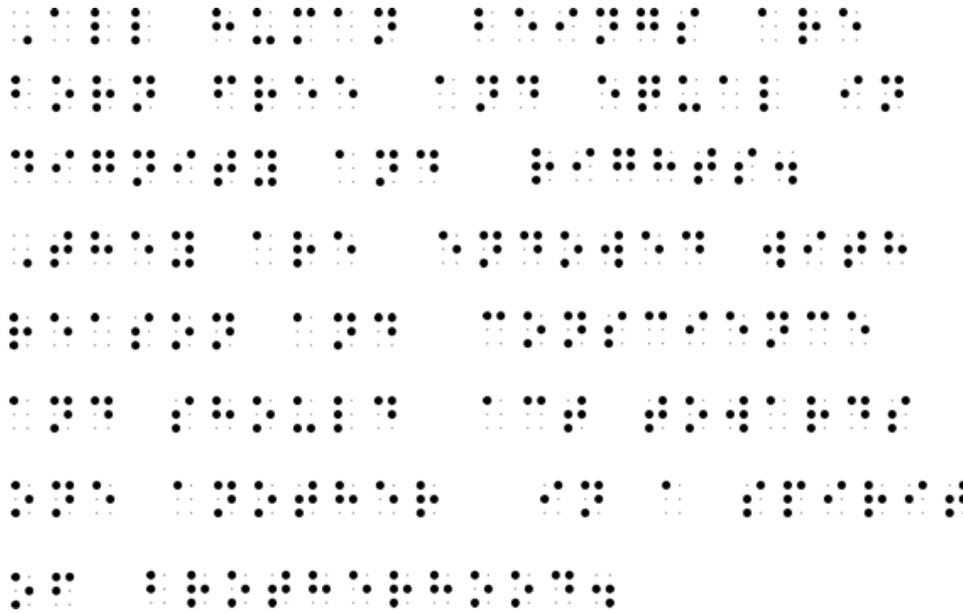
Figure 7. Article 1 of the universal declaration of human rights in Braille (Source: Anon, 1999a)

Figure 8. The moon alphabets (Source: Anon, 1999b)

The text in Figure 7 is transliterated as "All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood."

1.3.2 The Moon Alphabets

The Moon system of embossed reading was invented in 1845 by Dr William Moon of East Sussex. The Moon is a simple method of reading based upon the standard alphabet. The Moon alphabet is made up of 14 characters used at various angles, each with a clear bold outline. For many elderly blind people especially, Moon is easier than the more complex Braille system, although many people gain confidence from learning Moon to move onto Braille. The Moon alphabets are presented in Figure 8.

2. Materials and Methods
2.1 Signal Processing and Reading Machine Technologies

Signal processing is the extraction of information bearing attributes from measured data, and any subsequent transformation of those attributes for the purposes of detection, estimation, classification, or waveform synthesis. It is observed that the signals typically used in signal processing are functions of time, such as temperature measurements, velocity measurements, voltages, blood pressures, earth motion, and speech signals. Most of these signals are initially continuous signals (also called analogue signals) which are measured by sensors that convert energy to electricity. Some of the common types of sensors used for collecting data are microphones, which measure acoustic or sound data; seismometers, which measure earth motion; photocells, which measure light intensity;

optical scanners, which measure printed character representation; thermistors, which measure temperature; and oscilloscopes, which measure voltage.

When continuous electrical signals are collected from sensors, the continuous signal is converted to a digital signal (a sequence of values) with a piece of hardware called an analogue-to-digital (A/D) converter. Once digital signals are collected, computer could be applied to digital signal processing (DSP). These DSP techniques are designed to perform a number of operations such as: removing noise that is distorting the signal, extracting information from the signal, separating components of the signal, encoding the information in a more efficient way for transmission, detecting information in a signal just to mention a few of signal processing techniques. For some applications, an analog or continuous output signal is needed, and thus a digital-to-analogue (D/A) converter is used to convert the modified digital signal to a continuous signal. Another device called a transducer can be used to convert the continuous electrical signal to another form; for example, a speaker converts a continuous electrical signal to an acoustical signal.

In this section the three basic signals processing techniques for a reading machine are presented first from a theoretical point of view, secondly from an implementation point of view, and lastly from an applications point of view. The theoretical point of view includes the development of mathematical models and the development of software algorithms and computer simulations to evaluate and analyze the models both with simulated data and with real data. Real-time implementation can use VLSI (very large scale integration) techniques, with commercial DSP chips, or it can involve custom design of chips, MCMs (multichip modules), or ASICs (application-specific integrated circuits).

2.2 Mathematical Model: Fourier Transform

The Fourier transform is a mathematical tool that is used to expand signals into a spectrum of sinusoidal components to facilitate signal analysis and system performance. The Fourier transform is also used for spectral analysis, or for spectrum shaping that adjusts the relative contributions of different frequency components in the filtered result. In other applications the Fourier transform is important for its ability to decompose the input signal into uncorrelated components, so that signal processing can be more effectively implemented on the individual spectral components. Decorrelating properties of the Fourier transform are important in frequency domain adaptive filtering, sub band coding, image compression, and transform coding.

Classical Fourier methods such as the Fourier series and the Fourier integral are used for continuous-

time (CT) signals and systems, i.e., systems in which the signals are defined at all values of t on the continuum $-\infty < t < \infty$. A more recently developed set of discrete Fourier methods, including the discrete-time (DT) Fourier transform and the discrete Fourier transform (DFT), are extensions of basic Fourier concepts for DT signals and systems. A DT signal is defined only for integer values of n in the range $-\infty < t < \infty$. Fourier methods are particularly useful as a basis for digital signal processing (DSP) because it extends the theory of classical Fourier analysis to DT signals and leads to many effective algorithms that can be directly implemented on general computers or special-purpose DSP devices.

2.2.1 Classical Fourier Transform for CT Signals

The CT Fourier transform is useful in the analysis and design of CT systems, i.e., systems that process CT signals. Fourier analysis is particularly applicable to the design of CT filters which are characterized by Fourier magnitude and phase spectra,

i.e., by $|H(j\omega)|$ and arg. $H(j\omega)$, where $H(j\omega)$ is

commonly called the frequency response of the filter.

A CT signal $s(t)$ and its Fourier transform $S(j\omega)$ form a

transform pair that are related by the equation (1) for any $s(t)$ for which the integral (1a) converges (Madisetti and Williams, 1999):

$$s(j\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} \, dt \qquad (1a)$$

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} s(j\omega) e^{j\omega t} \, dt \qquad (1b)$$

Equation (1a) is simply called the Fourier transform, whereas Eq. (1b) is called the Fourier integral. The

relationship $S(j\omega) = F\{s(t)\}$ denotes the Fourier

transformation of s(t), where $F\{\}$ is a symbolic notation

for the integral operator and where ω is the continuous frequency variable expressed in radians per second. A transform pair s(t) $\leftrightarrow$ S(jω) represents a one-to-one invertible mapping as long as s(t) satisfies that condition which guarantee that the Fourier integral converges.

The operation of uniformly sampling a continuous time signal s(t) at every T sec is characterized by Eq. 2 presented below:

$$s_a(t) = \sum_{n=-\infty}^{\infty} s_a(t)\,\delta(t-nT)$$
$$= \sum_{n=-\infty}^{\infty} s_a(nT)\,\delta(t-nT) \qquad (2)$$

Where, $\delta(t)$ is a symbol used to denote a CT impulse function that is defined to be zero for all $t \neq 0$, undefined for $t = 0$, and has unit area when integrated over the range: $-\infty < t < \infty$. Since $s_a(t)$ is in fact a CT signal, it is appropriate to apply the CT Fourier transform to obtain an expression for the spectrum of the sampled signal:

$$F\{s_a(t)\} = F\{\sum_{n=-\infty}^{\infty} s_a(nT)\delta(t-nT)\}$$
$$= \sum_{n=-\infty}^{\infty} s_a(nT)[e^{j\omega T}]^{-n} \qquad (3)$$

Since the expression on the right-hand side of Eq. (3) is a function of $e^{j\omega T}$ it is customary to express the transform as F $(e^{j\omega T})$ = F $\{s_a(t)\}$. If ω is replaced with a normalized frequency, $\omega^/ = \omega / T$, so that $-\pi < \omega^/ < \pi$, then the right side of Eq. 3 becomes identical to the discrete time Fourier transform that is defined directly for the sequence s[n] = s$_a$(nT) (Madisetti and Williams, 1999).

2.2.2 DT Fourier Transform

The DT Fourier transfom (DTFT) is obtained directly in terms of the sequence samples s(n) by taking the relationship obtained in Eq. (3) to be the definition of the DTFT. By letting T = 1 so that the sampling period is removed from the equation and the frequency

variable is replaced with a normalized $\omega^/ = \omega$T, the DTFT pair is defined by Eq. (4). In order to simplify notation it is not customary to distinguish between ω and $\omega^/$, but rather to rely on the context of the discussion to determine whether ω refers to the normalized (T = 1) or to the unnormalized (T $\neq$ 1) frequency variable.

$$S\left(e^{j\omega^/}\right) = \sum_{n=-\infty}^{\infty} s[n]e^{-j\omega^/ n} \qquad (4a)$$

$$s[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} s(e^{j\omega^/})e^{jn\omega^/}d\omega^/ \qquad (4b)$$

The spectrum S (e$^{j\omega/}$) is periodic in $\omega^/$ with period 2π the fundamental period in the range $-\pi < \omega^/ < \pi$

sometimes referred to as the baseband, is the useful frequency range of the DT system because frequency components in this range can be represented unambiguously in sample form (without aliasing error). In much of the signal-processing literature the explicit primed notation is omitted from the frequency variable. However, when so many related Fourier concepts are discussed within the same framework.

By comparing (Madisetti and Williams, 1999) Eqs. (3) and (4a), and noting that $\omega^/ = \omega$T we see that:

$$F\{s_a(t)\} = DTFT\{s[n]\} \qquad (5)$$

8

Where,

$$s[n] = s(t)\big|_{t=nT}$$

This demonstrates that the spectrum of s_a (t) as calculated by the CT Fourier transform is identical to the spectrum of s[n] as calculated by the DTFT. Therefore, although s_a (t) and s[n] are quite different sampling models, they are equivalent in the sense that they have the same Fourier domain representation.

A reading machine relies on three basic technologies as follows:

- Scanner technology to scan an image into computer memory

- Digital image processing or optical character recognition (OCR) technology to convert the image to text

- Text-to-speech (TTS) technology to convert the text into intelligible speech.

3. Results

The reading machine, is the combination of three ubiquitous technologies, namely the scanner technology, the optical character recognition technology, and the text-to-speech technology.

3.1 Composite Parts of Reading Machine
3.1.1 Scanner

The scanner technology comprising the lamp, mirror, lens, CCD, and ADC converts the printed text to a bitmapped signal that is easily interpreted by the processor. The IRIS-Pen handheld scanner is the most fit for this work. It works just like a highlighter. Simply slide it over printed information from books, newspapers, magazines, faxes, letters, spreadsheets etc. and instantly it converts words and numbers into the reading processor application.

3.1.2 Processor

The processor, which engulfs the digital image processing technology and the text-to-speech technology, is the brain behind the entire operation. The output signal from the digital image processing is fed into the front-end compartment of the TTS. This first part has to process the signal through: text analysis, phonetic analysis, and prosodic analysis. A phoneme is

4. Discussion
4.1 Operation of Reading Machine

The sequence of operation of the reading machine begins when the scanner using an integrated scanning array scans the letters in each word and feeds

the sound associated with each letter. These signals in turn drive the speech synthesizer circuits in the back-end block compartment. Fig. 9 gives the block diagram representation of the reading machine.
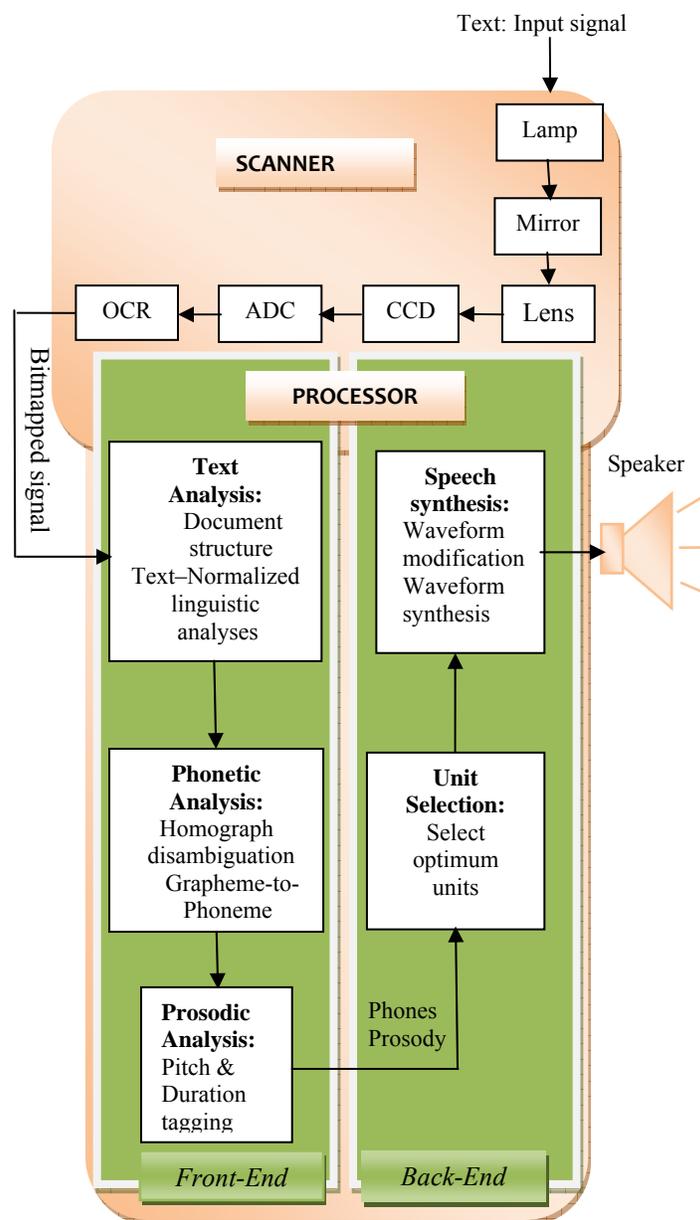

Figure 9. Block diagram of reading machine

the data directly into the processor. The Euraka handheld scanner and Iris-Pen express 6 handheld scanner are capable of doing this work perfectly. The Iris-Pen hand held scanner has these technical specifications: an universal serial bus (USB) interface, a personal computer platform, 16-bit (64 k colors)

maximum color depth and dimensions; width 1.41 inch, depth 0.94 inch, height 5 inch and weight 0.24 lb. Figure 10 shows samples.

Figure 10. Sample of Iris-Pen handheld
Scanners (Source: Anon, 2000)

A shape analysis program identifies the words and convert them into bitmap text code. If necessary, other programs using contextual and other clues assist in the identification. The TTS processes are made up of two giant blocks namely the front-end and back-end blocks.

4.1.1 Front-End Processing
The front-end section accepts text as input and produces a sequence of phones and associated prosody at its output. The front-end section can be subdivided into three distinct blocks: text analysis, phonetic analysis, and prosodic analysis.

The text analysis block performs a preprocessing step to analyze the document structure and organize the input sentences into manageable lists of words. In particular, punctuation must be correctly handled. For example, the text analysis block must understand that the colon in '23:45' indicates a time, and to disambiguate between an end of sentence period and decimal point such as in the sentence 'It is 3.14 miles to the city.' Text normalization deals with transforming abbreviations, acronyms, numbers, dates, and times into full text. This requires careful processing. For example, '20/08/1976' must be transformed into 'twentieth of August nineteen seventy six' and not erroneously as 'twenty forward slash zero eight forward slash one thousand nine hundred and seventy six'. It should be clear from these examples that the performance of the document structure and text normalization tasks is critical for ensuring accuracy of the TTS system. The text analysis block also performs some linguistic analysis. The part of speech category (e.g. noun, verb, adjective, etc.) for each word is determined based on its spelling.

The phonetic analysis block is concerned with grapheme-to-phoneme conversion (also called letter-to-sound conversion). Pronunciation dictionaries are employed at word level to provide the phonetic transcriptions. In order to keep the size of the dictionary manageable, words are generally restricted to morphemes. A set of morphophonemic rules is applied to determine how the phonetic transcription of a target word's morphemic constituents is modified when they are combined to form that word. Automatic grapheme-to-phoneme conversion based on rules is used for words not found in the dictionary as a fallback, though this approach is often error prone. The phonetic analysis block must also provide homographic disambiguation. For example 'how much produce do they produce?' Contextual information can aid in selecting the right pronunciation. A popular approach is to use a trained decision tree called a Classification and Regression Tree (CART) that captures the probabilities of specific conversions given the context.

The prosodic analysis block deals with determining how a sentence should be spoken in terms of melody, phrasing, rhythm, and accent locations – factors critical to ensure both intelligibility and naturalness of the resultant speech. From the perspective of the speech signal, prosody manifests as dynamic pitch changes, amplitude and duration of phones, and the presence of pauses.

4.1.2 Back-End Processing
The back-end stage of a concatenative TTS synthesizer consists of storing, selecting, and smoothly concatenating prerecorded segments of speech (units) in addition to modifying prosodic attributes such as pitch and duration of the segments i.e. subject to the target prosody supplied by the front-end. This section takes into account some of the key design questions such as: what unit of speech to use in the database, how the optimum speech units are chosen given phonetic and prosodic targets, how the speech signal segments are represented or encoded, and how prosodic modifications can be made to the speech units.

Different types of speech unit may be stored in the database of a concatenative TTS system. Obviously, whole words may be stored. However, whole word units are impractical for general TTS due to the prohibitively large number of words that would need to be recorded for sufficient coverage of a given language. Also, the lack of coarticulation at word boundaries results in unnatural sounding speech.

Modern speech synthesizers have evolved away from using databases with a single, 'ideal' diphone for a given context to databases containing thousands of examples of a specific diphone. By selecting the most suitable diphone example at runtime, and in many cases avoiding making quality-affecting prosodic adjustments

to the segment, significant improvements in the naturalness of the speech can be obtained.

4.2   Software Implementation

Speech synthesis markup language (SSML) is a standard, extensible markup language (XML-based), markup annotation for instructing speech synthesizers how to convert written language input into spoken language output employed by NaturalReader software. SSML is primarily intended to help application by controlling aspects of the speech output such as pronunciation, volume, pitch and rate. SSML can also express playback of prerecorded audio.

4.2.1 Document Structure

SSML documents are identified by the media type application/ssml+xml. Table 4 summarizes the elements and attributes defined in SSML. The basic structure of an SSML document is illustrated in Figure 11:

```
<?xml version="1.0" encoding="UTF-8"?>
<speak version="1.0"
        xmlns="http://www.w3.org/2001/10/s
        ynthesis"
        xmlns:xsi="http://www.w3.org/2001/
XMLSchema-instance"
        xsi:schemaLocation="http://www.w3.
org/2001/10/synthesis
```

```
        http://www.w3.org/TR/speec
        h-synthesis/synthesis.xsd"
        xml:lang="en-GB">
Hello world!
</speak>
```

Figure 11. Sample structure of SSML document (Burke, 2007)

All SSML documents include the root element <speak>. The version attribute indicates the version of SSML and is fixed at 1.0. The default namespace for the SSML <speak> element and its children is indicated by the xmlns attribute and is defined as http://www.w3.org/2001/10/synthesis. The xmlns:xsi attribute associates the namespace prefix of xsi to the namespace name http://www.w3.org/2001/XMLSchema-instance. The namespace prefix is defined since it is needed for the attribut, xsi:schemaLocation.

The xsi:schemaLocation attribute indicates the location of the schema to validate the SSML document against. The xml:lang attribute indicates the language for the document and optionally also indicates a country or other variation. The format for the xml:lang value follows the language tag syntax. Table 4 illustrates examples of language identifiers.

The <p> element and <s> element can be used explicitly to demarcate paragraphs and sentences.

Table 4 Elements and Attributes Defined in SSML

Elements	Attributes	Description
<Speak>	Version Xmlns Xml:lang xmlns:xsi xsi:schemaLocation xml:base	Root element for SSML documents.
<lexicon>	uri type	References an external pronunciation lexicon document
<p>	xml:lang	Explicitly demarcates a paragraph
<s>	xml:lang	Explicitly demarcates a sentence
<audio>	src	Inserts a recorded audio file.
<phoneme>	ph alphabet	Provides a phonemic/phonetic pronunciation for the contained text.
<sub>	alias	Provides acronym / abbreviation expansions.
<say-as>	interpret-as format detail	Used to indicate information on the type of text construct contained within the element.
<break>	time strength	Controls the pausing or other prosodic boundaries between words.
<emphasis>	level	Requests that the contained text be spoken with emphasis.

\<voice\>	xml:lang gender age variant name	Requests a change to the speaking voice
\<prosody\>	pitch contour range rate duration volume	Provides control of the pitch, speaking rate and volume of the speech output.
\<mark\>	name	Places a marker into the text/tag sequence.
\<meta\>	name http-equiv content	Contains metadata for the document
\<metadata\>	—	Contains metadata for the document

(Burke, 2007)

4.2.2 Interpreting Text

The \<say-as\> element is used to indicate information about the type of text construct contained within the element and to help specify the level of detail for rendering the contained text. Interpreting the contained text in different ways will typically result in a different pronunciation of the content (although a speech synthesizer is still required to pronounce the contained text in a manner consistent with how such content is normally produced for the language).

The \<say-as\> element has three attributes: interpret-as, format and detail. The format, and detail attributes are optional. The \<interpret-as\> attribute indicates the content type of the contained text construct, e.g. date to indicate a date, or telephone to indicate a telephone number. The optional format attribute provides further hints on the precise formatting of the contained text, e.g. a value of dmy could be used to indicate that a date should be spoken in the format of date, then month, then year. The optional detail attribute indicates the level of detail to be spoken although it is not defined for many interpret-as types.

In Figure 12 below are some common examples of \<say-as\>:

\<say-as   interpret-as="date" format="mdy"\>2/3/2006\</say-as\>
\<!-- Interpreted as 3rd of February 2006 --\>
\<say-as   interpret-as="time" format="hms24"\>01:59:59\</say-as\>
\<!-- Interpreted as 1 second before 2 o'clock in the morning --\>

Figure 12. Sample structure of prosodic interpretation. (Burke, 2007)

4.3   Power Management

Li-Ion batteries are leading edge battery technology and are an ideal selection for use on portable computers and cellular phones due to their high energy density and high voltage. A typical Li-Ion cell is rated at 3.6V and this is three times more than the typical NiCd or NiMH cell voltage (1.2V).

4.3.1 Features of Lithium Ion Batteries

These features are as follows:

- High energy density that reaches 400 Wh/L (volumetric energy density) or 160 Wh/Kg (mass energy density).
- High voltage. Nominal voltage of 3.6 V or even 3.7 V on newer Li-Ion batteries.
- No memory effect. Can be charged any time, but they are not as durable as NiMH and NiCd batteries.
- High charge currents (0.5-1A) that lead to small charging times (around 2-4 hours).
- Flat discharge voltage allowing the device to stable power throughout the discharge period.
- Typical charging Voltage 4.2 ± 0.05V.
- Charging method: constant current - constant voltage (CV-CC).
- Typical operation voltage 2.8 V to 4.2 V
- Recommended temperature range 0-4 ^{0}C

4.3.2 Safety Circuits inside the Li-Ion Battery Pack

Inside a Li-Ion battery pack there is always a safety circuit that consists of four main sections: the controller IC, control switches, temperature fuses, and the thermistor (Figure 13). The controller IC monitors each cell (or parallel cells) voltage and prevents the cells to overcharge or over discharge controlling accordingly the cutoff switches. Also the voltage across the switches is monitored in order to prevent over current. The control switches usually comprise FET structures that cutoff the charge or discharge depending on the control signals of the controller IC. The temperature fuses cutoff the current if the control switches experience abnormal heating. This fuse is not recoverable. The thermistor, usually called PTC measures the battery temperature inside the pack. Its terminals are connected to the charger so it can sense the temperature of the pack and control the charge current until the battery is fully charged.
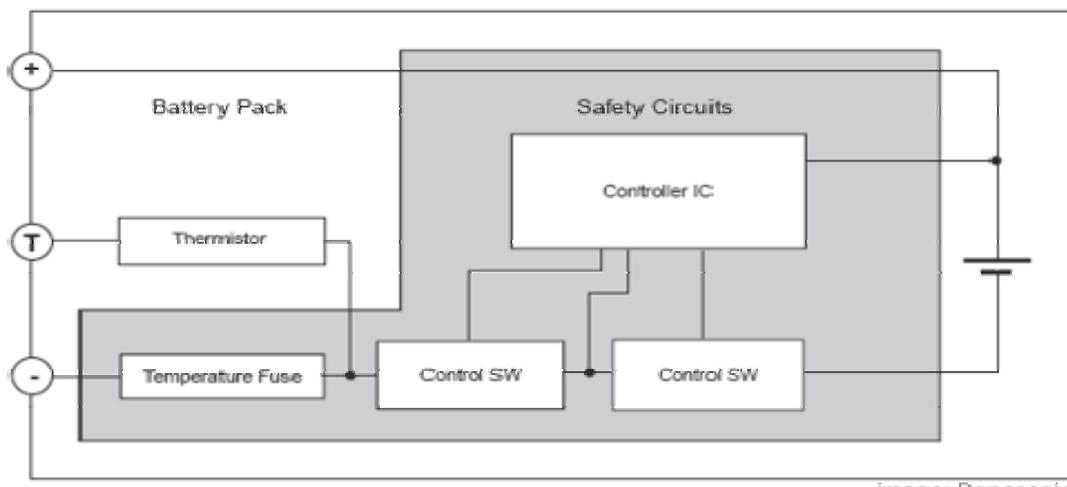

Figure 13. A typical block diagram of Li-Ion battery pack
(Source: Anon, 2004)

5. Conclusions

The architecture of a reading machine designed to achieve a high rate of correct interpretation of text by the blind and visually impaired has been presented. Three ubiquitous technologies were invoked: the scanner, the optical character recognition, and text-to-speech technologies. Multiple algorithms in a Fourier transform domain were used in signal and image processing. With the implementation of the reading machine developed, the feasibility of reading unconstrained printed materials will be achieved and information should be carried indiscriminately to the blind and visually impaired.

Acknowledgments

Correspondence to:
Normanyo Erwin
University of Mines and Technology
P. O. Box 237
Tarkwa 0362,
Western Region
Ghana

Telephone number: +233 (0)24 221 4103
Facsimile number: +233 362 20306
E-mail address: enormanyo@umat.edu.gh

References

[1] Anon, 1999a. Braille, retrieved on 6th February, 2008 from http://www.omniglot.com/writing/braille.htm.

[2] Anon, 1999b. Moon, retrieved on 6th February, 2008 from http://www.omniglot.com/writing/moon.htm.

[3] Anon, 2000. Scanners. Retrieved on 14th February, 2008 from http://www.irislink.com

[4] Anon, 2004. Li Ion construction, retrieved on 15th December, 2007, from http://www.electronicslab.com/articles/Li_Ion_reconstruct

[5] Burke, D., 2007. Speech Processing for IP Networks, John Wiley & Sons Ltd, The Atrium, Southern Gate, pp 79 – 88

[6] Dogbe, L., 2004. The Consequence of Being Blind in Ghana, retrieved on 6th January, 2008, from http://www.ghanaeyefoundation.org/index.htm

[7] Madisetti, V, Williams, D. B., 1999. (ed.) Digital Signal Processing Handbook. CRC Press, ghghghgh, 1st Edition, 1776 pp. ISBN: 084 9385725

[8] Microsoft Encarta Reference Library DVD, 2007

[9] WHO (World Health Organization), 2009a. Causes of Blindness and Visual Impairment. Retrieved on July 5, 2009 from **http://www.who.int/blindness/causes/en/**

[10] WHO (World Health Organization), 2009b. Magnitude of Blindness and Visual Impairment. Retrieved on July 5, 2009 from http://www.who.int/blindness/causes/magnitude/en/index.html

7/11/2009