A Handheld Colour-to-Speech Device for the Visually Impaired

Gabriel McMorrow
Zandar Technologies
Deans Grange, Dublin, Ireland
gmcmorrow@zandar.iol.ie

Paul F Whelan
Vision Systems Laboratory
School of Electronic Engineering
Dublin City University
Dublin 9, Ireland
whelanp@eeng.dcu.ie

Abstract

A hardware device is presented that converts colour to speech for use by the blind and visually impaired. The use of audio tones for transferring knowledge of colours identified to individuals was investigated but was discarded in favour of the use of direct speech. A unique colour-clustering algorithm was implemented using a hardware description language (VHDL), which in-turn was used to program an Altera Corporation's programmable logic device (PLD). The PLD maps all possible incoming colours into one of 24 colour names, and outputs an address to a speech device, which in-turn plays back one of 24 voice recorded colour names.

To the author's knowledge, there are only two such colour to speech systems available on the market. However, both are designed to operate at a distance of less than an inch from the surface whose colour is to be checked. The device presented here uses original front-end optics to increase the range of operation from less than an inch to sixteen feet and greater. Because of the increased range of operation, the device can not only be used for colour identification, but also as a navigation aid.

1 Using Audio Tones for Colour Recognition.

This project originally intended to use audio tones as the method to convey colour to a blind or visually impaired person. This original choice was partly due to the fact that the device was intended to be a small hand-held piece of hardware, and the use speech synthesis might complicate the design. The first method attempted was the writing of VHDL code to recognise one of 16 colours, by outputting a square wave whose frequency identified each colour, i.e. a colour-controlled oscillator. The square wave outputs were converted to sine waves with enough current to drive a small speaker. To make these audio tones more recognisable to the user, one might map these 16 tones to two octaves of the diatonic music scale. This would mean that a blind person could be trained to associate musical notes with colours by learning the musical scale. However, the use of two octaves restricts us to the use of only 16 colours. One possible method of increasing this range is to use the intensity of each tone. The intensity could be broken out into four different levels, i.e. dark red, red, bright red, pale red.

But how can we transfer this knowledge of intensity to the user without using more than 16 tones? One method would be to play these 16 tones, or notes, using four different musical instruments, i.e. middle 'C' on the cello might indicate "dark red", whilst middle 'C' on the "harmonica" might indicate "bright red", etc.

However, we now need to electronically synthesise four waveforms which would approximate four musical instruments, (note: they would only need to be discernible from each other as a 'cello' is from a 'harmonica'). *What makes middle 'C' on the 'cello' sound different from middle 'C' on the harmonica?* The answer is the number of harmonics of the fundamental frequency of middle ‘C’ generated by each instrument, coupled with the relative amplitudes of these harmonics.

Rather than electronically synthesise the sound of each instrument, a speech chip [ISD, 94] was used, see Section 5. This could also be used to record different instruments. We accept that there may be applications where the use of audio tones may be preferable to the spoken colour name, but we contend that in this particular application, the use of direct speech is preferable.

2 Colour Clustering

What is colour clustering and is it necessary? The answer to these questions will become intuitively clear to the reader, by examining the results of trying to identify colours without using a clustering technique. Before proceeding, however, let us examine some fundamental properties of colour. Every colour we see, is in fact a combination of the three primary colours, red, green and blue. The “colour cube” illustrated in Figure 1, is a graphical representation of this fact. The first step in colour identification involves splitting the incoming light into it’s red, green, and blue primary component intensities. In order to use digital circuitry to decide what the incoming colour is, these primary intensities need to be converted to analogue voltages and then digitised.

This naturally leads to the following question: *What should the resolution of the digitisation process be?*, i.e. how many bits should the analogue RGB components be digitised into. Suppose that 2-bit digitisation were used. This would provide the ability to discern 4 different shades of red, green, and blue. This in-turn implies we have the ability to identify 64 colours. If such a scheme were used, it would be necessary to give names to these colours. To aid in naming these colours, they first needed to be displayed. Once displayed, it became obvious that it was futile to try to give names to these colours without using some kind of quantitative method. For this purpose an RGB Colour Table [Lindley, 92] was used. The disadvantage of using this method would be the use of colour names not used in daily conversation, names such as “Cinnabar Green”, or “Medium Aqua Marine”. These are names that most sited people might have difficulty visualising in their minds, and are not suitable for describing colours to people blind since birth. To a certain extent this is a debatable point, however the next disadvantage is particularly critical.

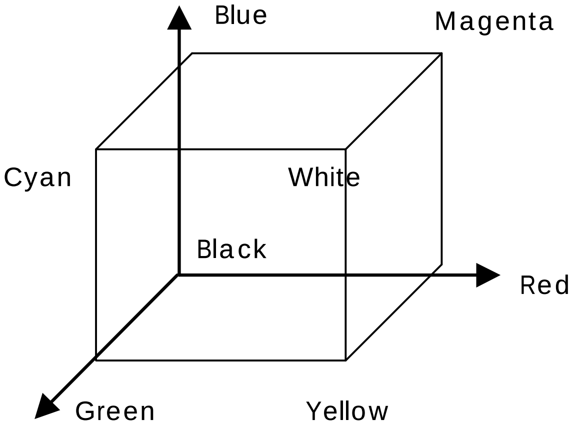

Figure 1. Colour cube.

The diagonal of the colour-cube shown in Figure 1, is a line in colour-space where the red, green, and blue components are equal to each other. This in fact is the greyscale, which extends from black, through dark grey, and finally to white. This diagonal line of greyscale is extremely sensitive, such that points slightly off the diagonal contain colour, not greyscale. When other colours other than greyscale are explored the error does not appear as serious, as it is difficult to tell, for example, close shades of green apart, however, for greyscale, small tints of colour are noticeable.

The only way to reduce this quantization error is by increasing the colour resolution, i.e. increasing the number of shades of the primary colours, red, green, and blue. It was found that 4-bit digitisation gave a visually acceptable colour quantization error for greyscale colour bins. However, using 4-bit digitisation provides the ability to discern 2^4 x 2^4 x 2^4 = 4096 colour bins. *Does this imply we need to derive 4096 different colour names?* The answer is no, because there are large regions or clusters of the same colours. For example, there are large regions of yellow, orange, green etc. This was observed when using MATLAB's Image Processing Toolbox to display colours found at the centroid of 4096 colour bins. Having the colours displayed enables the viewer to visually segregate or cluster regions of red, orange, yellow, etc. Figure 2 illustrates 3 of 16 layers of the colour cube digitised to 4-bit Red, Green, and Blue.

3 Design of Cluster Logic Using VHDL

As logic designs get larger, gate-level descriptions become unmanageable, making it necessary to describe designs in more abstract ways, and mandatory to adopt a higher-level, top-down design approach [Wang, 1994]. Logic diagrams and Boolean equations have been used as methods for hardware description, but the complexity of these methods increases rapidly as the system complexity increases. Hardware Description Languages (HDLs) evolved as a solution. An HDL is similar to a high-level software programming language and provides a means of:

1). Precise yet concise description of a design at desired level of detail.
2). Design capture for automatic design systems, e.g. high-level synthesis tools.
3). Incorporation of design changes and corresponding changes in documentation.
4). Design/user communication interface at the desired level of abstraction.

In VHDL, a hardware component is represented by the design entity, which in turn is a composite of an entity declaration and an architecture body. The entity declaration provides the "external" view of the component's ports, which are the points through which data flows into the device from the outside world. This includes the component's ports, which are the points through which data flows into and out of the component [Lipsett, 1992]. The entity declaration for the clustering device is as follows:

```
ENTITY  cluster IS
    PORT( red: IN bit_vector(3 downto 0);
          green: IN bit_vector(3 downto 0);
          blue: IN bit_vector(3 downto 0);
          color: OUT bit_vector(9 downto 0)
     );
END cluster;
```

The architecture body provides the "internal view"; it describes the behaviour or the structure of the component. The following VHDL is a shortened version of the architecture body used to define the behaviour of the cluster device, and is almost a direct transcription of the visual clustering shown in Figure 2.

```
ARCHITECTURE arch1 OF cluster IS
BEGIN combin1: PROCESS(red, green, blue)
               CASE  blue IS   -- cluster the first 6 layers of the RGB color cube
WHEN "0000" | "0001" | "0010" | "0011" | "0100" | "0101" =>
IF green <= "0011" THEN IF red <= "0011" THEN color <= "0100110111";
ELSIF ( ( red = blue) and (red = green) )   THEN color <= "0101000010";
ELSE color <= "0100101010";
END IF;    etc. etc .
END PROCESS; END arch1;
```

The architecture body has a process sensitivity list. This means the device is inactive, and only becomes active when it senses a change of the input signals red, green, and blue. The entire architecture essentially makes decisions of the form: if we measure X % of blue, Y % of red and Z % of green, then the colour is referred to as pink.

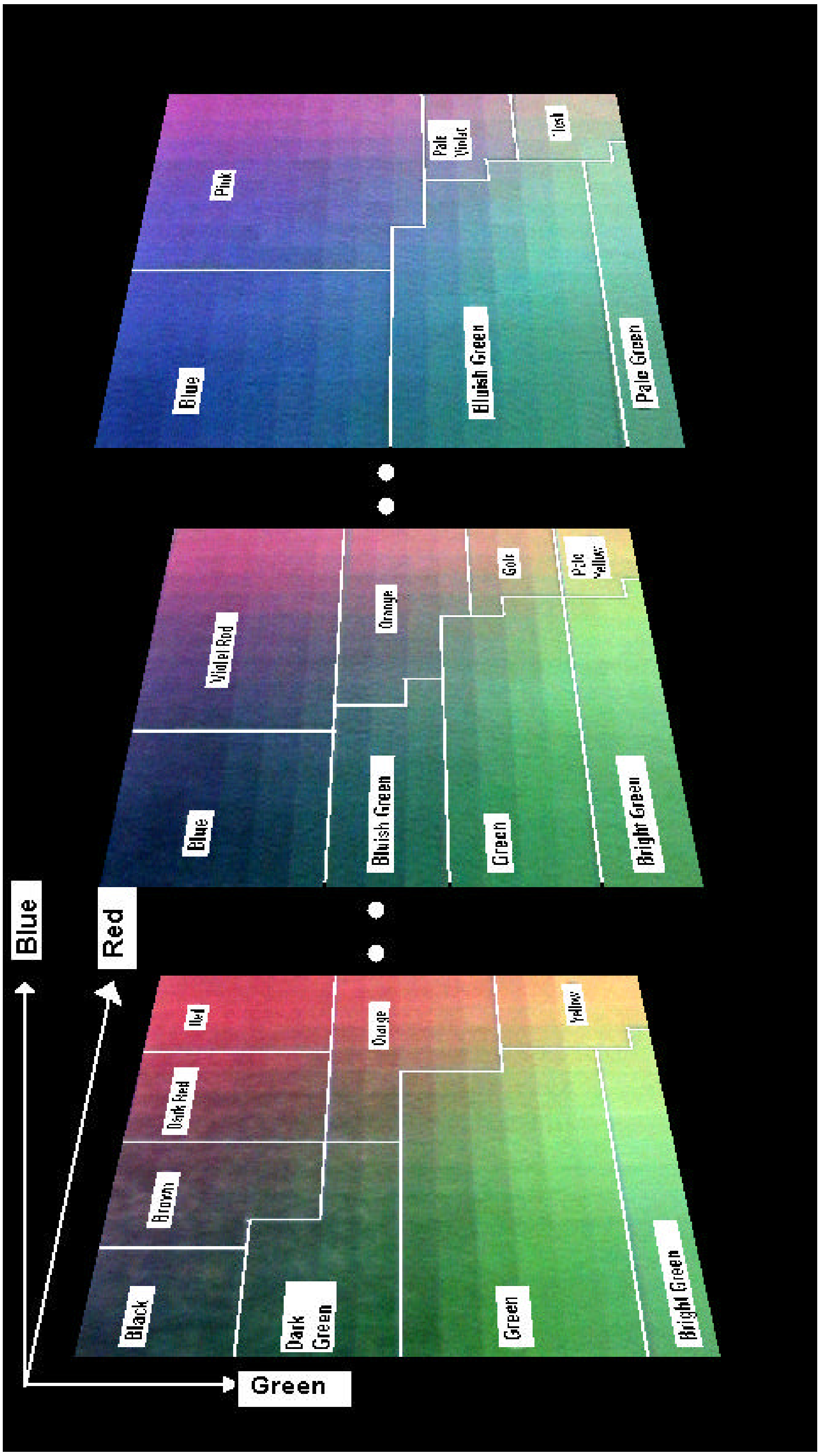

Figure 2. Three of 16 layers of the colour cube digitised to 4-bit Red, Green, and Blue.

4 Front End Optics and Sensors

Of the two colour-to-speech conversion systems available on the market at the time of writing, both need to be operated within an inch of the surface whose colour is to be checked. A main goal of this research was to design a device, which would operate over much larger ranges. In order for the device to operate at greater distance ranges, it is necessary to use a lens to gather the incoming light, and project it onto the light sensitive surface of the photodiodes. In addition, the lens is required to have a narrow field of view (FOV). The narrow field of view is important so that the colour of an object's surface is tested at a minimal surface area. This is to avoid the identified colour being an averaged colour, i.e. the result of checking a large FOV encompassing multiple surfaces of different colours.

The disadvantage of using a fixed single lens is that the minimal area being checked grows in proportion to the distance the surface being checked is from the lens. Therefore, using a single fixed lens implies that the continuity of colour of the surface in question determines the range the device must be operated at to get a correct colour reading. For example, pointing the device at a cloudless sky should return blue because although the minimal surface area being checked is large, the sky should be a constant blue over all the area being checked. Similarly, the green grass in a field should return green when the device operates at large ranges. However, for surfaces whose colour changes over areas of, say, an inch diameter, the device should be within 8 feet of the surface, to get correct colour readings, etc. Figure 3 illustrates how a plano-convex lens can be used to limit a detector's FOV [Edmund, 95]. The angle β subtends half the field of view of the lens, and is given by:

$$b = \frac{D\,2}{2 * f} \quad radians$$

Where *D2* is the diameter of the photodiode used to convert the light to current, and *f* is the focal length of the lens. Placing the detector one focal length behind the lens ensures focus of objects one focal length in front of the lens. It was found that the reduction in FOV resulted in a single incoming colour at a time, scanning over an average sized room.

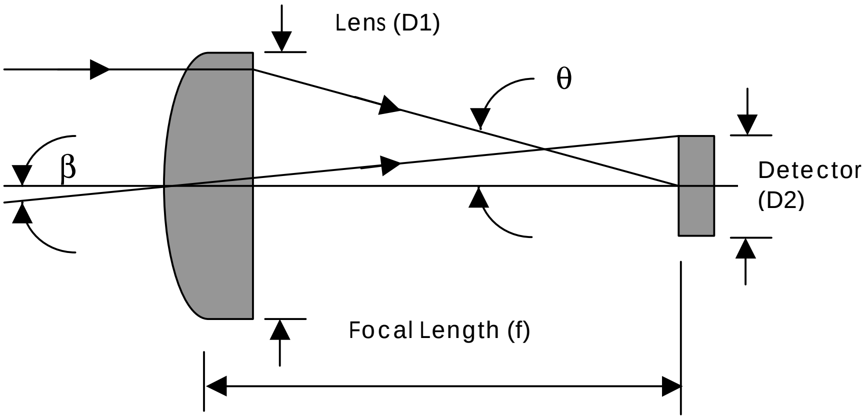

Figure 3. Using a plano-convex lens to limit a detector's FOV.

Obviously it is important to centre the photosensor on the axial line of the lens. Knowing where the light will go is, however, only the first step in the design of a light projection system; it is just as important to know how much light is transmitted. Figure 3 illustrates how the numerical aperture controls the angle of light accepted by the lens. As the F-number decreases, θ (the acceptance angle) increases and the lens becomes capable of gathering more light. In the design of any lens system it is important to consider the term throughput (TP), a quantitative measure of transmitted light energy. Because the photodiode is an area and not a point, lens diameter affects throughput even when the F-number remains constant. Because the device is to be physically small, a small lens diameter is required whilst at the same time maintaining a large throughput. For example for a particular lens system the TP's for each of three lens choices are: TP(6mm lens) = 16.77, TP(12mm lens) = 19.04, and TP(25mm lens) = 19.69. In this example the 12mm diameter lens is the best compromise between diameter and throughput. The following table is a subset of calculations that resulted in a final choice of the plano-convex lens used in this application:

Focal Length F	Lens Outer Diam (mm) D1	½ FOV degrees β^o	TP	Object Area Diam Inches at 1 ft.	Object Area Diam Inches at 2 ft.	Object Area Diam Inches at 3 ft.	Object Area Diam Inches at 4 ft.	Object Area Diam Inches at 8 ft.	Object Area Diam Inches at 16 ft.
75	18	1.03	0.255	0.216	0.432	0.647	0.863	1.72	3.45
100	19.8	0.77	0.175	0.161	0.323	0.484	0.645	1.29	2.58
100	25	0.77	0.277	0.161	0.323	0.484	0.645	1.29	2.58

A photo-diode whose response matches closely the human eye was chosen. The device is the BPW 21 photodiode, whose main application is exposure and colour measurement. It comes in a hermetically sealed case, has a radiant sensitive area of 7.5 mm^2, and its typical sensitivity is 7 nA/lux. Typical room light is about 700 lux. Before considering the amplification used with these photodiodes, the important topic of colour filtering needs to be addressed. The incoming light needs to be split up into its primary colour components, (red, green, and blue), prior to being converted to electrical signals. Various colour filter options are available. Two choices of design for colour separation was examined, see Figure 4.

Choice A is the better choice as only a single lens is used whose diameter may be large. Choice A is made possible by the availability of colour separating Dichroic filters. A light filtering system using a 45 degree blue reflector/corrector, 45 degree red reflector/corrector, and green corrector set allow a single source of light to be split into it's component parts. An aluminium structure was designed and manufactured to hold the glass filters at their appropriate angles, and support the three photodiodes.

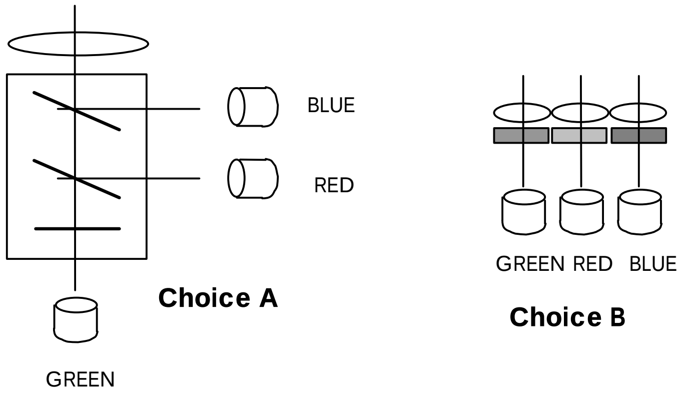

Figure 4. Colour separation design choices.

5 Front-end Analogue Hardware Design

5.1 Analogue to Digital Conversion

Assuming that the incoming primary component intensities have already been converted to analogue voltages, it is then necessary to convert the RGB analogue signals to 4-bit binary logic levels. This will form the appropriate input into the cluster device described previously. The A/D device chosen was Analog Devices' AD7575, which uses the successive approximation conversion technique. This provides a necessary high conversion speed of 5 microseconds. A standard 555 timer was used to control the start of conversion and the reading of data from the AD7575.

5.2 Light Sensor Amplification

Next in consideration is the amplification of the individual red, green, and blue signals coming from the photodiodes. It is desired that the A/D converters output 4-bit binary patterns spanning the range of 0 to 15. To accomplish this it is necessary for the output signals of the photodiodes to span the analogue voltage range 0 to 2Vref, i.e. 0 to 2.4V. A single supply TLC271 operational amplifier was were chosen for amplification.

6 The ISD2590 Single-Chip Voice Record/Playback Device

The ISD2590 [ISD, 94], manufactured by Information Storage Devices provides an excellent method for verbal description of identified colours required in this project. It allows cueing of recorded colour names, which can then be directly addressed. Therefore, once the incoming red, green and blue intensities have been digitised, and clustering performed by the Altera Programmable Logic Device [Altera , 94], one of 24 colours (addresses) can be input to the speech chip to cause it to "speak" out the name of the colour. The ISD2590 allows a total of 90 seconds recording time. It provides excellent voice quality, and the two output pins provide direct speaker drive capability of about 12.5 mW RMS (25 mW peak) into a 16 Ω speaker. This is enough to be clearly heard from the other side of an average room. The ISD2590 provides a total of 90 seconds of recording time, and a total of 600 individual messages. This gives rise to the term message resolution, which is 90/600 = 150 milliseconds. The address bits control the message start pointer (MSP). The value of an address to access a particular message x seconds into the recording is given as follows: *Address (in decimal) = (x seconds) / (message resolution)*, i.e. to access a message 1.5 seconds into the recording use address = 1.5 / .15 = 10. To access a message 3.0 seconds into the recording use address = 3.0/. 15 = 20, etc. These addresses, 10, 20, 30 etc. require that each message be exactly 1.5 seconds long. It was for this reason that it was decided to use the Windows Audio Processing package COOL. COOL provides a variety of editing features such as amplification and message length adjustment of the digitally sampled and stored recording. This allowed the calculation of the addresses needed to be presented to the speech chip to access each of 24 different colour names, for example:

COLOUR NAMES	**Address (binary) into chip**	**Address (Decimal) into chip**	**Time in seconds into the total recording**
DARK RED	"0000000000"	0	0
RED	"0000001110"	14	2.1
VIOLET RED	"0000011100"	28	4.2
PINK	"0000100111"	39	5.85
Etc.	Etc.	Etc.	Etc.

7 Conclusions

At the time of concluding this report we were still awaiting delivery of the red 45-degree reflector colour filter. This restricted the full testing of the device. However, by manipulating light falling on the unfiltered red sensor, various intensities of RED could be achieved. When the device was presented with full blue, full red, and no green, pressing a button on the speech chip resulted in it "playing back" the word MAGENTA. Full red only generated the word RED, etc. This demonstrated correct operation of the A/D converters, the cluster logic and the speech chip. By tying the inputs of the A/D converters together, the greyscale was tested using a variable DC level in lieu of the sensors. For 0V DC the device said BLACK. When the DC level was increased it said DARK GRAY, GRAY, and finally WHITE at DC level 2.4V. This pseudo testing of the overall device brought to light many areas of improvement. The gain of the sensors is definitely a calibration issue. Lens choice provides a trade-off between shorter focal length (larger FOV) versus lens diameter to provide larger throughput (TP). These trade-off and calibration issues require further field testing of the device.

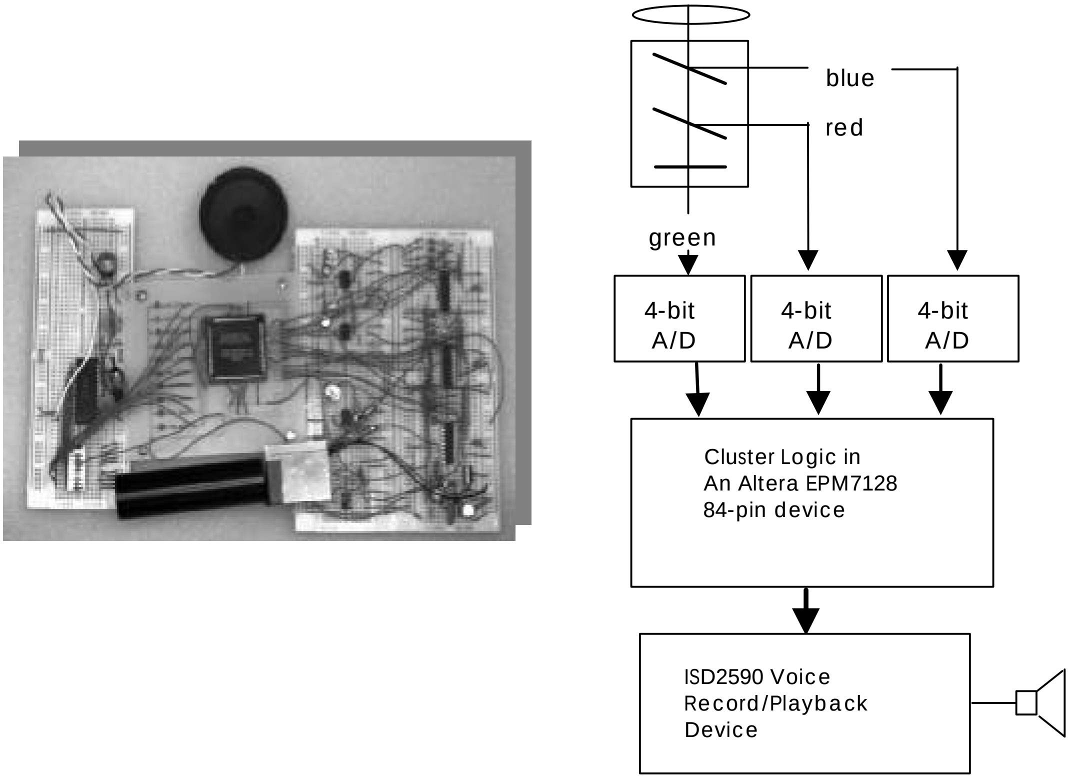

Figure 5. Block diagram of complete system, and photograph of prototype.

References

[Altera, 94], Altera Corp. (1994), "MAX+PLUS II Getting Started".

[Edmund , 94], Edmund Scientific (1994), *Annual Reference Catalog for Optics, Science, and Education*.

[ISD, 94], Information Storage Devices'(1994), *"Application Notes and Design Manual for ISD's Single-Chip Voice Record/Playback Devices"*

[Lindley, 1992], Lindley, C. A. (1992). *Practical Ray Tracing in C,* Wiley.

[Lipsett, 1992], Lipsett, R. (1989) *VHDL : Hardware Description and Design,* Kluwer Academic Publishers.

[Wang, 1994], Wang, X. (1994). *VHDL and High-Level Logic Synthesis,* Course Notes, School of Electronic Engineering, Dublin City University.

Gabriel McMorrow & Paul F Whelan (1997), "A Handheld Colour-to-Speech Device for the Visually Impaired", Proc. *IMVIP'97 - Irish Machine Vision and Image Processing Conference 1997*, 10 - 13 September 1997, University of Ulster, Northern Ireland, pp ??-??.