CONTROL OF A SIMPLE DC MOTOR ROBOT EQUIPPED WITH ULTRASONIC SENSORS VIA A FIELD PROGRAMMABLE GATE ARRAY AND A SPEECH RECOGNITION BOARD AND MICROPHONE

Andrew Jason Tickle [*,1,3] **Paul Kenneth Harvey** [*,2,4]
James Ross Buckle [*,1,5] **Jeremy Simon Smith** [*,6]

[*] *Intelligence Engineering and Industrial Automation Research Group, Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool, L69 3GJ, U.K.*

Abstract: This paper documents a feasibility study showing the initial development into building a practical robotic speech controlled system; this can be used to assist people with disabilities, gaining more of their independence back. Speech recognition is a very useful, and novel approach for controlling devices; the system presented here uses the 48 pin voice recognition integrated circuit, HM2007 package to perform the actual recognition for either speaker dependent or independent systems. Also described is how the Field Programmable Gate Array (FPGA) system is interfaced with the HM2007. Coding for the latter was done via the use of our Hardware Description Language, called VHDL. An extensive Altera vector waveform analysis of the systems is shown to verify that the control and safety systems function as they were designed to. Also shown is how the FPGA is used to control the H-bridge driver chip, rather than design a system from scratch, this is due to the fact that the number of lines required would make building a logic circuit very time consuming and very complex. There is also a detailed view of how noise affects the control mechanism, how the safety features are built into the system to avoid errors and accidents from occurring, and from the system being misused.

Keywords: Speech Recognition, Field Programmable Gate Array (FPGA), HM2007 Speech Board, Object Avoidance, Speech Filter

[1] This work was supported in part by the EPSRC (Engineering and Physical Science Research Council).

[2] This work was supported in part by a UK Student Recruitment Office Scholarship (UKSRO).

[3] Andrew J. Tickle is a Ph.D. student with the Univ. of Liverpool - Corresponding author - Tel: 0044 1517944602; Fax: 0044 1517944540; Email: a.j.tickle@liv.ac.uk

[4] Paul K. Harvey was an M.Sc student and is now a Ph.D. student with the Univ. of Liverpool - Email: pkh@liv.ac.uk

[5] James R. Buckle was a Ph.D. student with the Univ. of Liverpool and is now a Research Assistant at the Univ. of Glasgow - Email: james.buckle@gmail.com

[6] Jeremy S. Smith is a Professor of Electrical Engineering with the Univ. of Liverpool - Email: j.s.smith@liv.ac.uk

[7] The authors would also like to thank Dr Andrew Shaw and Professor Armed Al-Shamma'a for their comments on this work.

[8] The authors would also like to thank all the people who took part in the speech recognition survey and those whom spoke for the Mandarin language tests for providing there voices for use in this project.

1. INTRODUCTION

Speech is an ideal method for robotic control and communication, the system investigated here exploits device inter-connectivity, the computationally intensive task of the speech recognition is retained on a separate custom Application Specific Integrated Circuit (ASIC), while the robots control mechanism and all other functions are retained by the FPGA. Another advantage to this approach is its programmability, the HM2007 can be programmed and trained with the unique words that the user wishes, again good for this application as only a few specific words are in use. The speech recognition can then be easily interfaced with the robot. The HM2007 is extremely flexible, easily re-trainable, and can support multiple languages or dialects with either speech dependent or independent recognition (see section 3.1) meaning that no extra coding is required for the FPGA.

The concept of speech recognition for control purposes is not new, and work in recent years has progressed to obtaining an effective speech recognition system on an FPGA, based around Hidden Markov Models (Nedevschi et al., 2005), due to the increasing potential of parallelization and pipelining in speech coding. This work has been motivated to assist elderly and disabled people to maintain as much of their independence as possible. It is a well known fact that these people do not like relying on care workers for various aspects of their daily life, this work will start some of the basic technology that could help these people keep their dignity and some more of their freedom.

The implementation will involve connecting the FPGA to the HM2007 via the latter's output display pins. Control signals are produced from the FPGA based on the signals coming from the HM2007, and these are designed for use with a H-bridge driver chip, which would then provide the motion and control of the motors. Other considerations include safety systems, built into the system to avoid harm to the user. The two layers of protection include an ultra-sonic sensor net, that will stop or steer the robot out of the way of oncoming objects in the path. The second is a series of emergency stop switches, placed at strategic locations, and another switch on the user hand rest as an emergency stop.

2. THE SIMPLE ROBOT

The selected geometry was the most common non-holonomic system in robotics, the coaxial drive (McComb, 2002), this is the most basic form and consists of two independent driven wheels mounted on either side of the robot. It is referred to as coaxial due to the fact that the wheels share a common axis, though in practice, many robots do not but wheelchairs do. One or two un-powered omnidirectional castors are placed at the center line over the center of gravity at the front to support the base (they can be placed at the front and back). Example of such systems are Nomad Scout, SmartRob and the EPFL robots, Pygmalian and Alice. A "car-type" steering system is another way to control the robot, this method was investigated but this type of control method is not as manouverable as coaxially steered robots, examples of this type of robot are Neptune (Carnegie Mellon University) and Hero-1. This was proved when the robot was put through a basic a obstacle course, it had to complete the course in the minimum amount of time whilst being controlled by the user, and avoiding the obstacles placed in there path.

3. SPEECH RECOGNITION AND CONTROL COMMANDS

3.1 Speech Recognition Board

Speech recognition is classified into two categories, speaker dependent and speaker independent: *Speaker dependent* systems are trained by the individual, who will be using the system and achieve a 95 percent accuracy. The drawback to this approach, is that the system only responds accurately to the individual who trained the system. *Speaker independent* is a system trained to respond to a word, regardless of who speaks and therefore the system must respond to a large variety of speech patterns, inflections and enunciation's of the target word, and as a result recognition accuracy is lower. The HM2007 can be set up to identify words or phrases 1.92 seconds in length at the cost of reducing the word recognition vocabulary number to 20, this is a very useful package and has already had several practical applications (Rockland et al., 1998).

The speech recognition circuit for the HM2007 can be obtained from the company's website, Images SI Inc (Images SI). In manual mode, the HM2007 uses a simple keypad and digital display to communicate with and program the chip. When first initialized, the circuit checks the static RAM and if errors do not occur the board displays "00" and an indicator to show the system is "Ready" and waiting for a command. To train the circuit, the user presses the word number they want to train on the keypad, followed by speaking the word for recognition into the microphone. The basic words programmed here are "Forward", "Reverse", "Left", "Right", "Spin" and "Stop". The circuit is continually listening, when a trained word is spoken into the microphone, the number of the word is shown on the digital display.

3.2 The Speech Feasibility Study

In this section, there is an investigation into how accurate the HM2007 is at recognizing other people, with different accents from different nations, and different regions from these nations using a fixed template based on a single individuals own voice for to see how the system reacts. This is so that a gauge of how the speaker dependent system acts, since you don't want a wheelchair driving down the street and someone who is not the operator shouting "Left Left Left" and the wheelchair driving into the road because of a practical joker! Age is also a factor, it is known that from childhood to adulthood the voice of a male changes the most, but does it change to the point where it would affect recognition accuracy? The undergraduate population provided what have been classified here as the "teenage" samples, whilst the postgraduate and staff provided the "adult" samples. The test was were speakers said one of the control commands randomly into the microphone in sets of three, within an environment with either no background noise, normal "chatter", loud music and within a reverberation chamber, to see how the latter affects the system. Under each of these conditions the speaker will say the word in their normal voice, a whispered voice, a shout or loud voice and a slowed voice. The results of the UK participants are shown in Figures 1, where the symbols N/Q meant Normal voice in Quiet Background etc and UKAM meant United Kingdom Adult Male.

From the survey results there are troughs in the graphs, when the speaker was asked to shout or put on a slowed voice, this can be explained by the fact that its very hard to imitate someone with a slowed voice, such as if they are drunk for example, secondly, everyone's voice sounds different when they shout due to the differing high frequency components. The voice type results came to be similar, with the only major drop in recognition taking place between UK Teenage Males, the other main categories had similar responses and so that main template is very good. A similar trend was spotted in the other nationalities investigated apart from the case was reversed with Chinese Adult Males and Chinese Teenage Males. For the reverberation tests, it was discovered that reverberation produced better results. The reflected waves seem to reinforce the spoken command, and help to raise the recognition rate of the speaker using the system.

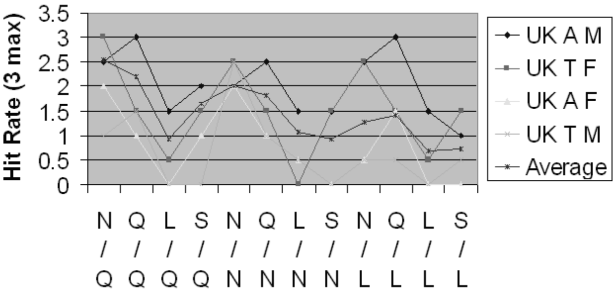

Fig. 1. United Kingdom participants survey results

Table 2: Logic inputs to activate the L298 chip

Function	Inputs (EnA EnB In1 In2 In3 In4)
Forward	111010
Reverse	110101
Left	010010
Right	101000
Spin	111001

3.3 Language Independence

In order to show that the same methods and processes can be applied to languages other than English, the system was also tested with commands spoken in Mandarin, the official language of the Peoples Republic of China. The test consisted of the same spoken commands as in English, saved into the HM2007 and was tested using native Chinese speakers and British who can speak Mandarin. The test subjects were selected from the student population from all levels of their degree at the University of Liverpool. Using a single word system, the system response was extremely similar to that of English, with the overall response slightly lower, but better than English in some areas when background noise of certain types was used.

3.4 Interfacing the FPGA with the HM2007 for Motor Control

The FPGA will process the incoming signal from the HM2007 and then output them so that they can activate the L298 Dual Bridge Driver chip (L298 Datasheet). Table 2 shows the truth table to get the L298 to perform the different movement operations such as "Forward" etc that would have to be produced by the output pins of the FPGA when certain words are spoken. However, the output voltages from the HM2007 only reach 3.6V, which means that they are not high enough to trigger the 5V connectors on *certain* FPGA boards, so the lower threshold pins are used when it comes to making the system work on any FPGA system.

4. VHDL CODED SYSTEM

4.1 The Safety Block

This checks if any of the safety switches have been activated, and if so it does not let the input signals from the HM2007 be transferred to the output, so that they can be processed by the hardware interface block. If none of the switches are activated, then the inputs signals simply pass through the block and come out the other side unaffected. The safety switch bank consists of the three primary switches in this test case, although more can be added at a later date if so desired. These are all connected in series, so if any of the switches are activated, there is a drop in voltage across the input pin to the FPGA, so the block then knows to cut the power to the motors thus stopping the robot from moving.

The ultrasonic transducers resonate at 40kHz, if this resonant frequency varies too much ($\pm 750Hz$), the performance of the transducers degrades rapidly. The ultrasonic receiver section was based around a 741 operational amplifier, while the ultrasonic transmitter was based around a CMOS 555 timer set, up in astable mode. The sensors are tuned to specific distances from the robot body, so that they generate a certain voltage when triggered, this is done by manipulation of the sensors range, by adjusting the corresponding resistor values in the circuitry (McComb, 2001). After tuning the transmitter, the circuits comparator circuit needed to be set up correctly, this was done via placing a solid object at varying distances, and readjusting the resistance in the comparator until a test LED just lit up. The comparators need to be set up so that when $V^- > V^+$ you get a zero output, and when $V^+ > V^-$ you get a positive output. These ultrasonic sensors would also be activated by anything in the robots path so the robot can take the action to avoid it. The output from the comparator gives an output signal either $+5V$ or $0V$ (for lower voltages a simple potential divider circuit can be used).

4.2 The Decoder Block

The decoder block is to allow the user to see the numbers of the commands being issued that would have been displayed on the two seven segment decoders on the HM2007, however since these are in use to connect to the FPGA board, the two onboard seven segment displays on the board can be used to display the current number being used by the HM2007. This lead to the development of a block that could decode the binary values being input to the board and lighting the corresponding seven segment displays on the board. The block checks all the input combinations in a look up table, and when it comes across the correct values for the input signals, it outputs the correct combination for the that particular input.

4.3 The Interface Block

The interface block, is the most challenging of all as it is required for the actual control of the L298, using the commands from the HM2007. The hardware interface works on the principal, that if one of the lines is activated, it sends out a code to the L298 telling it to move in the desired direction. The direct interfaces read the binary values and the values from the ultrasonic sensor comparators and uses them to determine which direction to move in and outputs the correct signal. If the command is “Reverse”, the robot will simply reverse, if in any other direction however, the ultrasonic sensors come into play. If the command is “Forward” and no sensors are triggered the robot will continue to move forward. If an ultrasonic sensor is triggered, however the robot moves to the side opposite to which the sensor is triggered. Whereas if the command is “Left” for example, the robot will only turn right if the left sensor is tripped and ignores the right sensor. The opposite is true for the “Right” directional command.

4.4 The Filter Block

From the speech survey, the recognition took a decline when noise was introduced into the system, and so a filter was designed on the FPGA tuned to the human vocal frequency from $300Hz$ to $3kHz$, so that only the users voice and background noise, and other high frequency components are removed. The signal goes into the FPGA, is filtered and then outputted to the input of the HM2007. Figure 2 shows a MATLAB simulation of the filter characteristics obtained from the filter design toolbox within the MATLAB environment, and which was used to create the VHDL filter.

This filter is considered separate, as it has no direct control over the motors, its purpose is to improve recognition to the control circuitry. The digital filter hardware for this application consists of the FPGA itself (which is based around the Altera APEX 20KE200 FPGA), and two SPI-compatible interface ICs: a 12-bit ADC and an optional 8-bit DAC. The digital filter is a rectangular-windowed band-pass finite impulse response (FIR) filter, implemented in VHDL. The design parameters have been selected to provide maximum performance in terms of filter frequency response, but most importantly minimum resource usage in terms of the number of Configurable Logic Blocks (CLBs) consumed. This particular FPGA hardware had

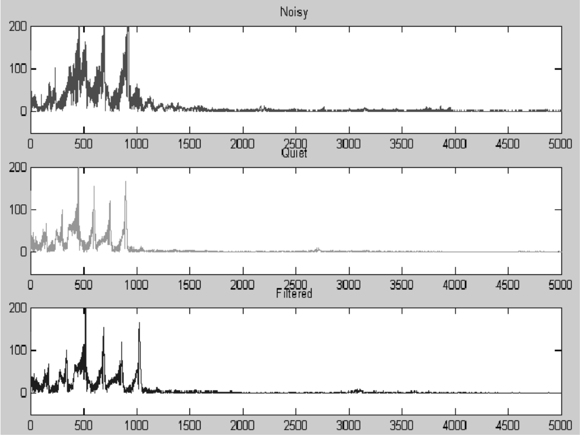

Fig. 2. MATLAB filter simulation using filter coefficients on a saved speech command in a (a) noisy environment, (b) quiet environment and the (c) filtered version

a fixed clock frequency of 33 MHz, from which the sampling clock was derived. The sampling frequency of 48 kHz was chosen specially to provide high definition in the lower frequency bands, and to make it easy to construct a very simple low-quality anti-aliasing analogue filter at the input. The 48 kHz 8-bit single-channel audio can be directly interfaced to stock computer audio hardware in the same unmodified native format which is similar to digital audio tape (DAT) or both consumer and professional format digital audio links. Samples are triggered at regular intervals by a simple hardware counter implemented in code, allowing 688 cycles per sample of the master clock for computation. The filter constants were imported from MATLAB's filter design tool to pass frequencies from 300 Hz to 3 kHz, this is shown in Figure 3. Frequencies outside the band roll off to -40 dB in 0.1 decades before the lower, and after the upper pass bands. With these filter parameters, the optimum filter order was determined to be of an order of 341, producing 342 constants. To minimise FPGA resource usage, the filter code consists of one combinational multiplier (one operation per cycle), which sequentially computes the output sample by looping through each filter constant stored in read-only memory and multiplying it with the corresponding delayed sample, this is shown in Figure 4. Due to an inefficiency of the design, two cycles per multiply-and-accumulate are required, leaving only 344 of the 688 computing cycles per sample to finish processing a sample. This can easily be corrected by performing register latching and multiplication sequentially in one clock cycle, rather than alternating between multiplying and accumulating. With 342 constants (or 684 constants with the corrected design), this leaves only 4 redundant cycles per sample, making the design up to 99.418 percent efficient in terms of clock cycle efficiency.

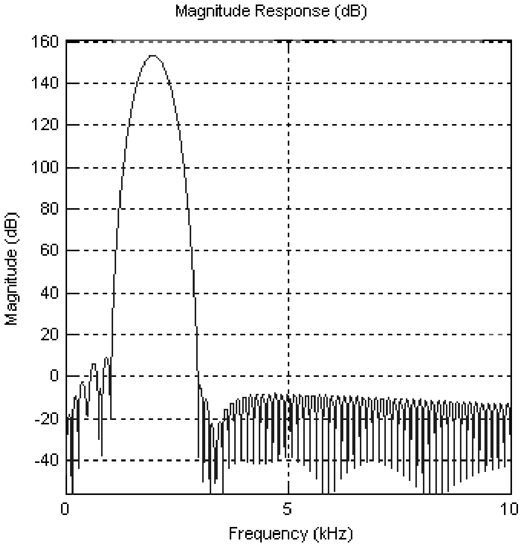

Fig. 3. Magnitude frequency response from the filter design toolbox in MATLAB

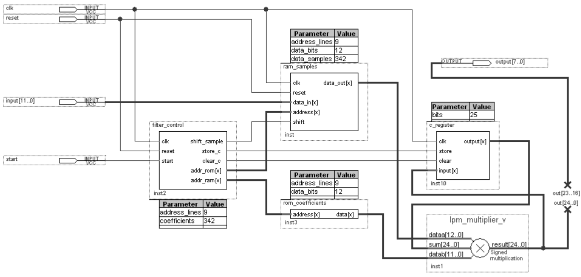

Fig. 4. VHDL filter system shown in the Quartus II environment

5. RESULTS

5.1 The Complete System Results

Due to the sheer number of simulations, it is *impossible to show* all the vector waveform analysis graphs here, and so the complete system simulations will be discussed here. The full VHDL system consists of all the previous blocks connected together, and simulated as a whole before they were downloaded to the FPGA for practical tests in the lab. This ensured that propagation delays and other time delays do not affect the performance of the system as a whole. Figure 5 shows the overall VHDL blocks in graphical, form inside the Quartus II environment and Figure 6 shows the simulation results obtained from the entire system. The area highlighted on the left shows that the Least Significant Bits (LSB) are unaffected by the ultrasonic sensors or switches, and still display current number of issued command. The square dotted area at the top left, shows that when the safety switches or ultrasonic sensors are activated, they override the given commands, and stops the robot when needed. The long dash dot area in the top right, shows that in the break between safety checks the output issued out is given by the correct code. Other simulations which are not shown checked that the system input counts up the values on the seven segment display change to the correct values, to display the correct numbers, when this number gets too high, it goes out of the range and so no useful output is given and the default value on the display is given, and that the output values to the L298 are correct.

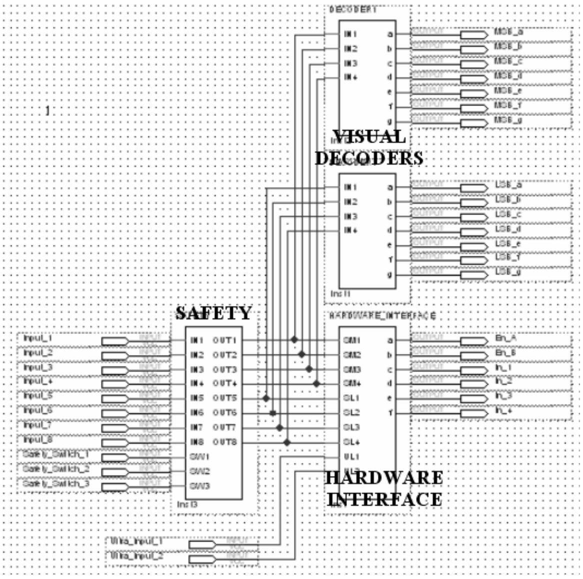

Fig. 5. Full control system shown in the Quartus II environment

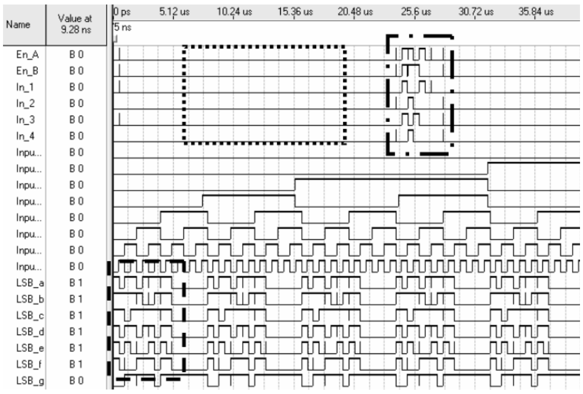

Fig. 6. Quartus II vector waveform analysis of the full control system

6. CONCLUSIONS

Here was presented the design of a low cost real-time speech controlled robotic system, the first phase in the development of a robotic wheelchair for the elderly or for a voice activated exploration robot that could be combined with a radio and camera for tele-presence multitasking applications. The evaluation of the design included both real-time implementation of the design on an APEX model FPGA, and a detailed resource allocation for more advanced features to be added on later models of Altera boards. The system tested using two languages: English and Mandarin. Also the details of each block where described and how each block fitted into the overall system. Extensive simulation results, show that the proposed structures function within their design parameters, and that there were no problems concerning timing delays caused by the coding.

From completing an FPGA resource allocations for each of the blocks, as the system only used between 26 and 69 percent of board resources, a large amount of free space will allow future work to allow 'Dead Reckoning' to be employed, where the users most visited locations can be stored as mathematical measurements in terms of the number of turns of the wheels and distances, and can be commanded to drive to a given location from a know start point by saying either "One", "Two", "Three" etc. The system can also be put through some more rigorous worst case situation simulations, to see how the system reacts when motors or sensors fail, and modify the code according to reduce the risk of harm to humans. The voice of the user, also changes under stress or excitement, this was initially planned to be a part of the speech survey, but the survey participants experienced difficulty when trying to simulate a stressed or panicked situation, and so this test was canceled and should be investigated in the future, so the end user would be able to operate such a device in a crisis situation such as a fire.

REFERENCES

Holmes J and Holmes W 2001 *Speech Recognition and Synthesis* 1st Edition pp 2 - 158 ISBN-10: 0748408575

Images SI Inc Website Speech Recognition - http://www.imagesco.com

L298 Dual Bridge Driver chip Data Sheet - http://www.st.com/stonline/books/pdf/docs/1773.pdf

McComb G, 2001 *Robot Builders Bonanza* McGraw-Hill Education pp 190 - 210 ISBN-10: 0071362967

McComb G 2002 *Constructing Robot Bases (Robot DNA)* McGraw-Hill Education pp 232 - 243 ISBN-10: 0071408525

Nedevschi S, Patra R A and Brewer E A 2005 *Hardware Speech Recognition for User Interfaces in Low Cost, Low Power Devices*

Rockland R H and Reisman S 1998 *Voice Activated Wheelchair Controller* IEEE Bioengineering Conference pp 128 - 129