

Project Report

March, 2008



S.I.S. (**S**peech **I**nteractive **S**ystem)

Phanikumar V V

&

Sagar Geete

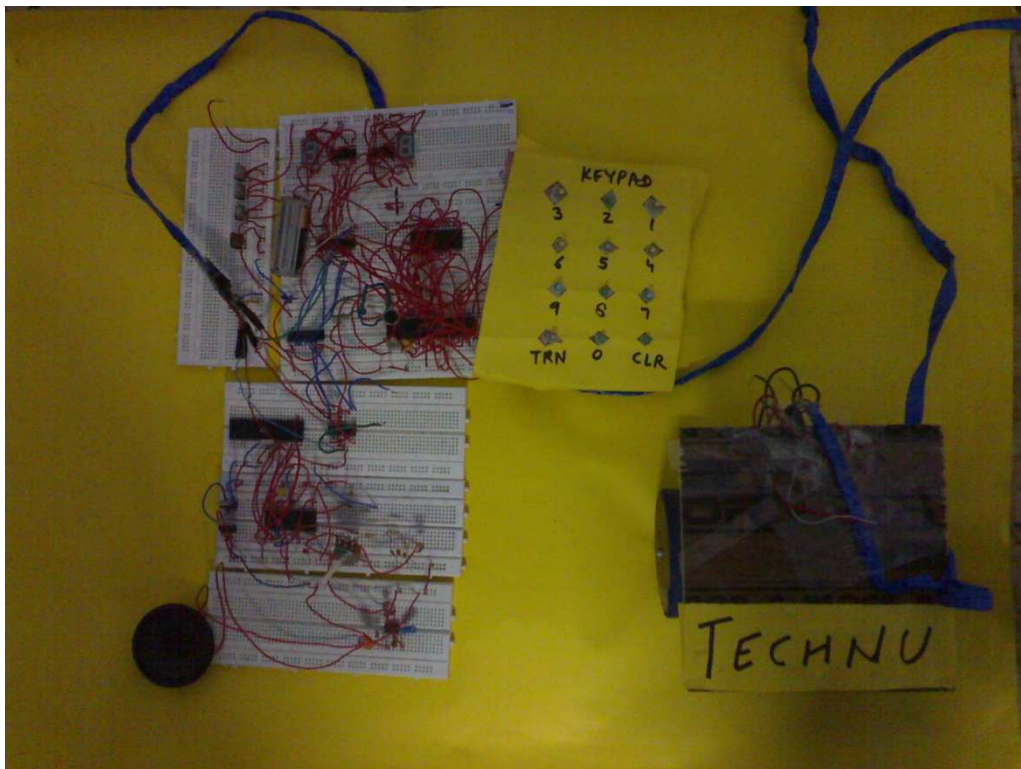
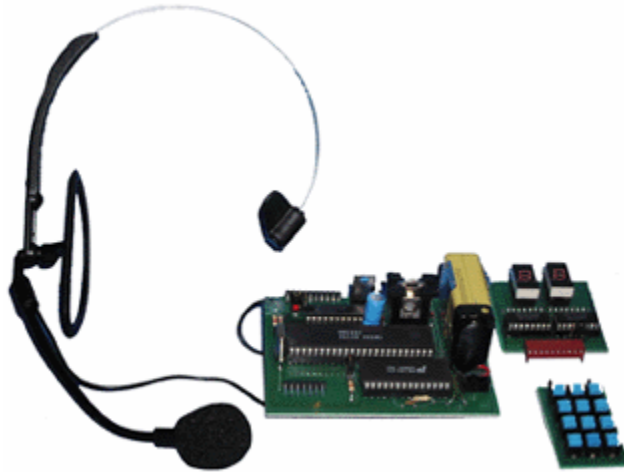
(Mining 2010, IT-BHU)

CONTENTS

1. Introduction	3
2. Utility	4
2. Speech Recognition	6
3. Speech Synthesizer	8
4. Speech Interactive System	10
5. Future Considerations	11

INTRODUCTION

Human Interaction with ROBOTS AND ELECTRONIC GADGETS using SPEECH is the basic motto of the S.I.S. (Speech Interactive System). The Voice Interface System (VIS) was at the heart of the project. The VIS consists of the voice response and speech recognition subsystems. The Speech interactive System is a complete easy to build programmable speech recognition & synthesis circuit. Programmable, in the sense that you train the words (or vocal utterances) you want the circuit to recognize. The synthesis of unlimited number of English words can made using the circuit.



UTILITY

In the near future, speech recognition will become the method of choice for controlling appliances, toys, tools, computers and robotics. There is a huge commercial market waiting for this technology to mature.

This project details the construction and building of a stand alone trainable speech recognition circuit that may be interfaced to control just about anything electrical, such as; appliances, robots, test instruments, VCRs TVs, etc. The circuit is trained (programmed) to recognized words you want it to recognize.

To control and command an appliance (computer, VCR, TV security system, etc.) by speaking to it, will make it easier, while increasing the efficiency and effectiveness of working with that device.

At its most basic level speech recognition allows the user to perform parallel tasks, (i.e. hands and eyes are busy elsewhere) while continuing to work with the computer or appliance.

Applications

- Command and control of appliances and equipment
- Telephone assistance systems
- Data entry
- Speech controlled toys
- Speech and voice recognition security systems

Software Approach

Currently most speech recognition systems available today are programs that use personal computers. The add-on programs operate continuously in the background of the computers operating system (windows, OS/2, etc.). These programs require the computer to be equipped with a compatible sound card. The disadvantage in this approach is the necessity of a computer. While these speech programs are impressive, it is not economically viable for manufacturers to add full blown computer systems to control a washing machine or VCR. At best the programs add to the processing required of the computer's CPU. There is a noticeable slow down in the operation and function of the computer when voice recognition is enabled.

Example of circuit implementation

The Talking Toaster

It's 3:00am. You're hungry. You've been up all night implementing a threads package for your Operating Systems course project. You stumble into the kitchen. Can you really be troubled with setting the toaster's heat setting, or activating the toaster's heating coils?

Of course not! That's where the Talking Toaster comes in. Instead of fiddling with the toast-quality dial or hitting the down level, the toaster will actually *ask you* for the settings. Even better, you can simply respond by speaking your reply -- no buttons to push, dials to spin, or lights to watch.

The operating instructions for the toaster are quite simple. When you want toast, ask the toaster for some toaster:

You: Toast.

The toaster will then ask you what your preferred toasting level is:

Toaster: How light?

Respond with light, medium, or dark.

You: Medium.

The toaster will then lower its bread tray, engaging the heating coils:

Toaster: Using setting medium. Lowering...

When the temperature has reached the desired threshold, the toaster raises the bread tray and disengages the heating coils:

Toaster: Raising... done!

That's all there is to it. Isn't that cool?!? Not only does the toaster *talk to you*, but you can talk to the toaster, and it **understands** you!

The technologies that the toaster would employ include:

- Speech recognition, using the HM2007 speech chip.
- Speech Synthesizer SPO256 chip.
- Microcontroller system control, via ATMEGA16. Appendix of the Final Report is the microcontroller code.
- An old toaster.
- Servo motor.
- Plain and simple ingenious engineering.

Speech Recognition

The speech recognition subsystem was built using Hualon Microelectronics Corporation's HM2007 speech recognition chip. This chip allowed for words to be recognized from a vocabulary of up to 40 1-second long words. The vocabulary was stored on an external 8K

SRAM. This SRAM was not powered by a battery, this removed a design constraint that otherwise would have hampered the project.

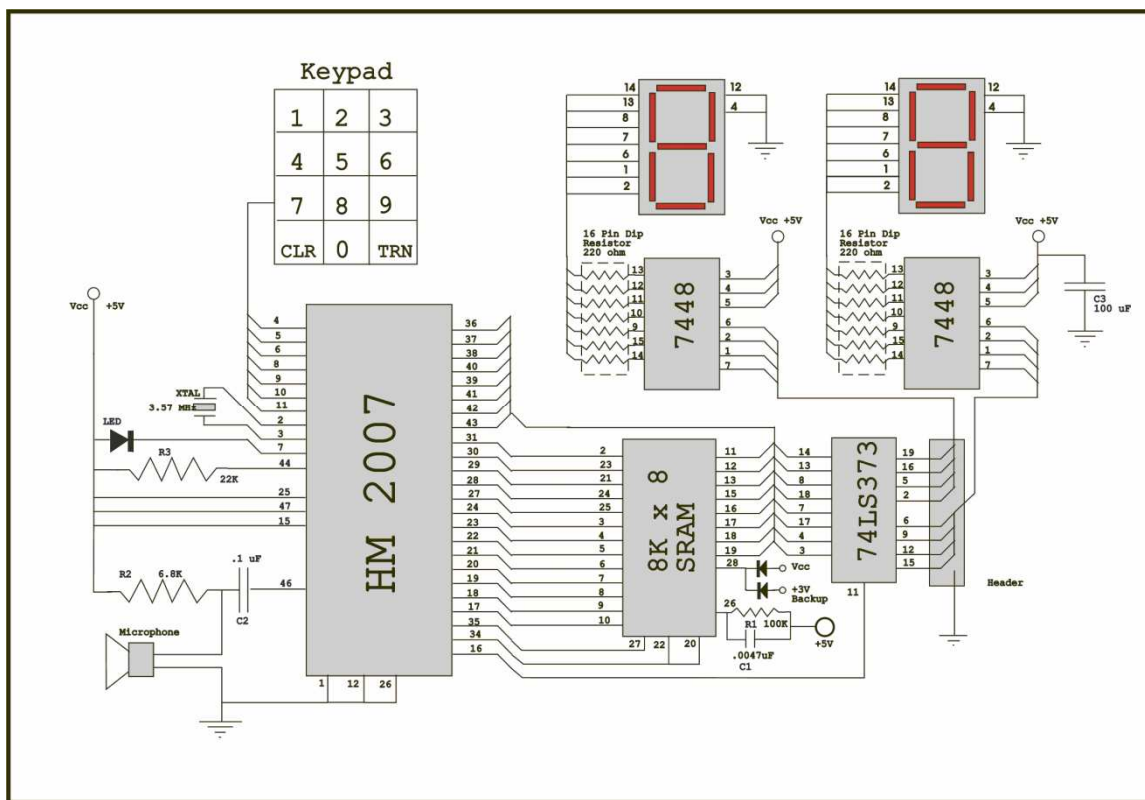
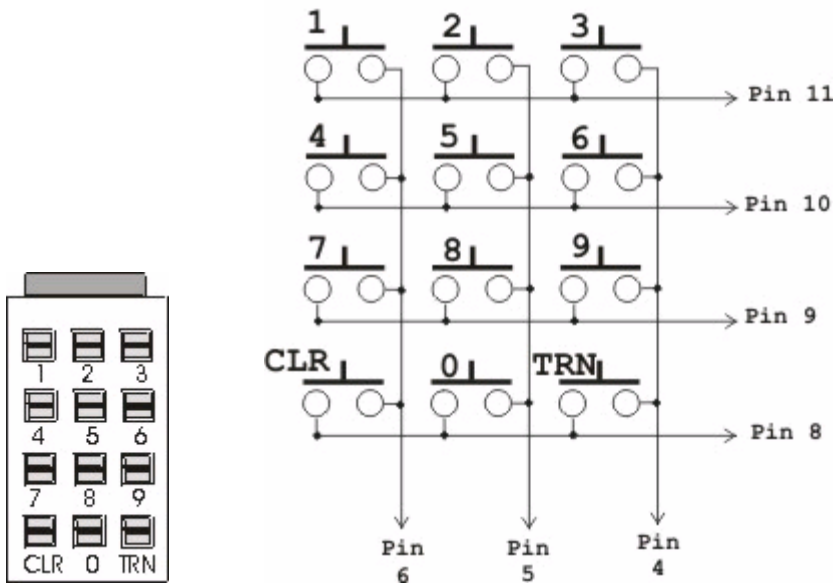
To use the speech recognition system of the HM2007, the user must train their voice prints on the chip. In the current version of Project code, the user is instructed on how to do this when the circuit is first plugged in. For each word that is to be recognized, the microcontroller asks the user to speak that word. Because the user may say the word differently (i.e., with slightly different inflections, etc.), the user is asked to say the word more than once (usually three times). For each time the user says the word, the HM2007 integrates this word into a neural network (this network is stored in the off-chip SRAM). Later, in recognition mode, the HM2007 tries to match the spoken word against other words in its neural net. If a match is made, the index of that word in the vocabulary is returned. If no match is found, or if the user spoke too quickly or too slowly, an appropriate error code is returned.

Thus, the HM2007 does not recognize a spoken word as an actual word, but rather as sounding like a word that it knows about. The HM2007 has no a priori knowledge of what the word 'back' should sound like.

The implementation of the speech recognition system was by far the most difficult engineering feat in the entire project. The chip came with a data sheet, but the information contained there in was in many instances unclear and even incorrect. Fortunately, after two weeks of intensive experimenting and designing, we had an understanding of the basic program flow required to train and recognize words.

The HM2007 is a 48-pin PDIP chip. The HM2007L has 4 I/O ports, a microphone system, and several control pins. To communicate with the 8K SRAM, there is a 13-bit address bus and a 8-bit data bus (two of the four I/O ports), as well as a memory read/write pin and a memory enable pin. To communicate with the keypad, there is a 4-bit wide K-bus, used for passing data to the HM2007, and a 3-bit S-bus, used for sending commands to the HM2007 (mainly commands to control the meaning of the K-bus).

Keypad

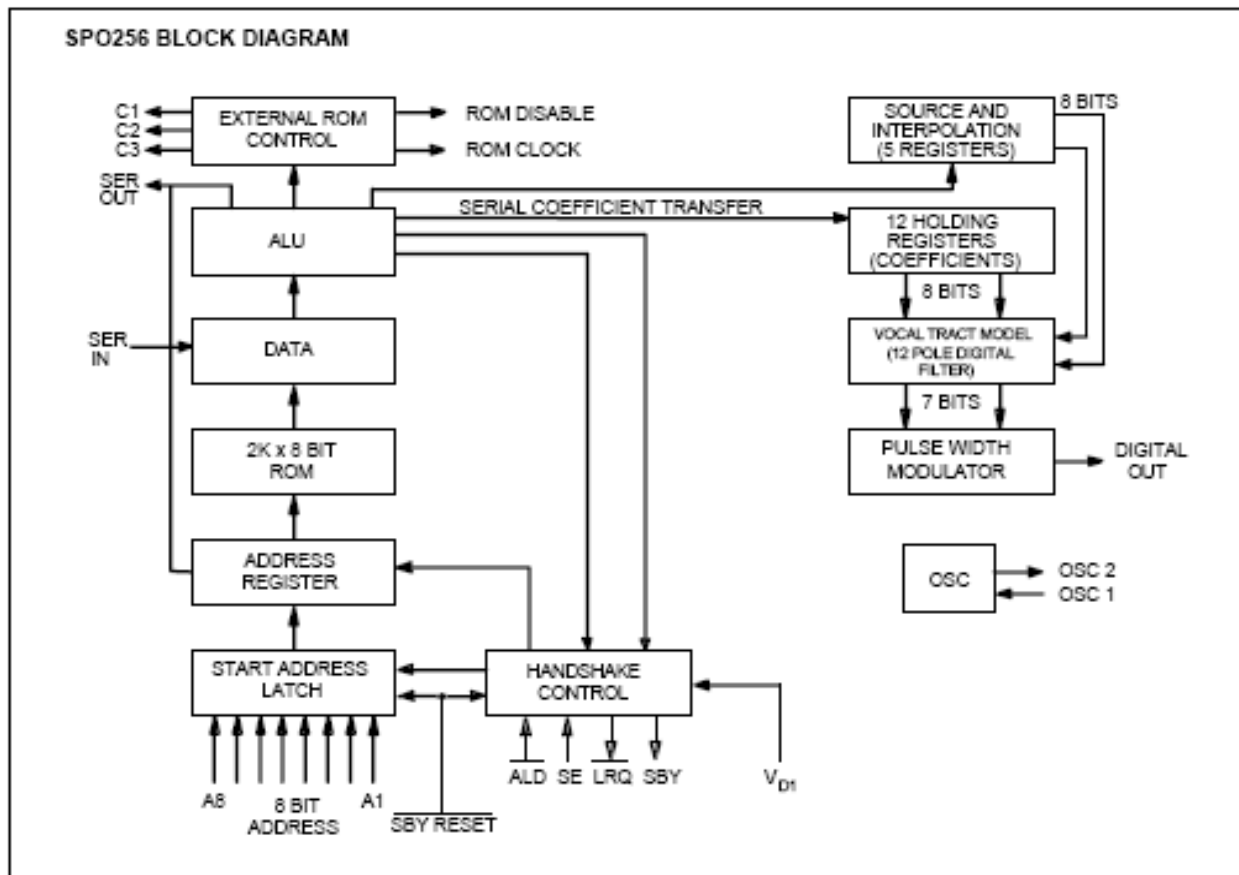


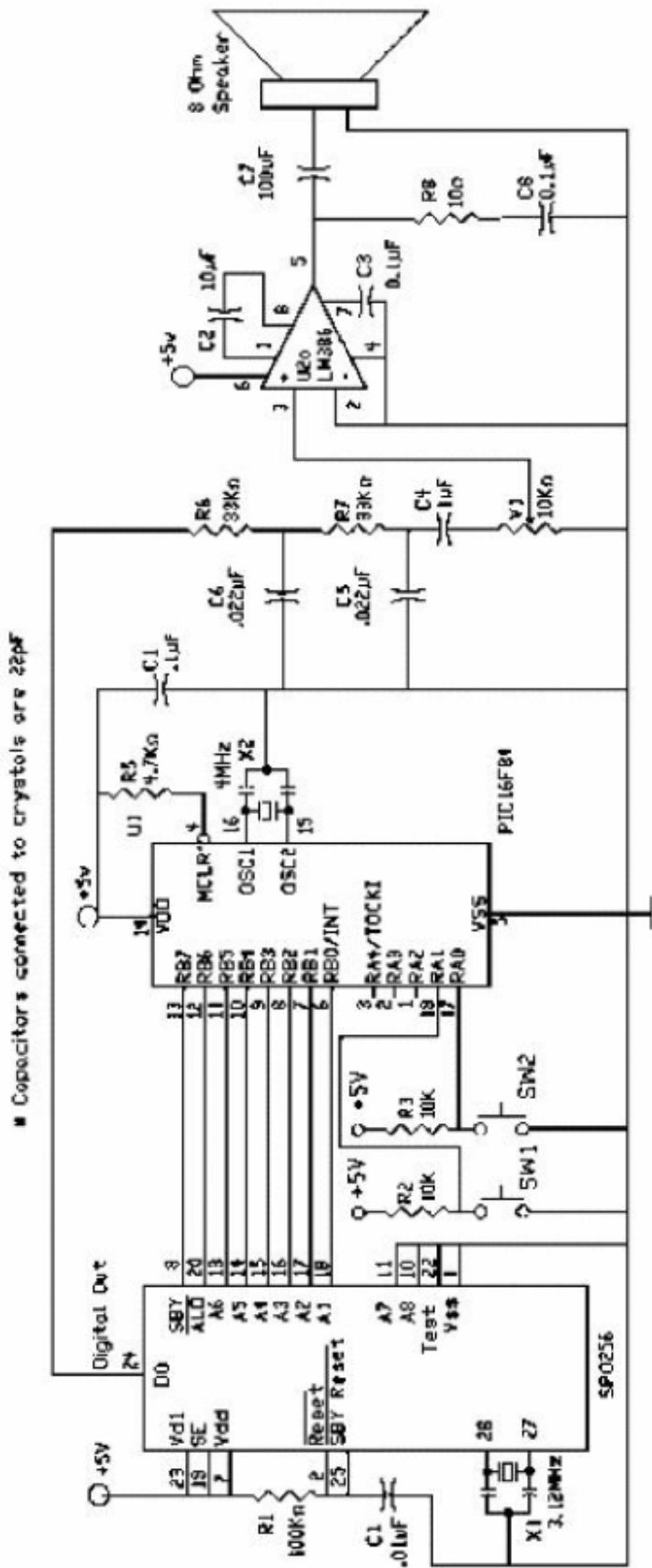
Speech Synthesizer

The voice response subsystem was implemented with an **SPO256 NARRATOR SPEECH PROCESSOR**. The SPO256 (Speech Processor) is a single chip N-Channel MOS LSI device that is able, using its stored program, to synthesize speech or complex sounds. The achievable output is equivalent to a flat frequency response ranging from 0 to 5 kHz, a dynamic range of 42dB, and a signal to noise ratio of approximately 35dB.

The SP0256 incorporates four basic functions:

- A software programmable digital filter that can be made to model a VOCALTRACT.
- A 16K ROM which stores both data and Instructions (THE PROGRAM).
- A MICROCONTROLLER which controls the data flow from the ROM to the digital filter, the assembly of the “word strings” necessary for linking speech elements together, and the amplitude and pitch information to excite the digital filter.
- A PULSE WIDTH MODULATOR that creates a digital output which is converted to an analog signal when filtered by an external low pass filter.





SPEECH INTERACTIVE SYSTEM

Interaction of speech recognition with speech synthesis is like mile stone in our project; because both recognition & synthesis looks independently less effective. But when we interfaced both with one another its applicability increases suddenly. We can feel actual part of humanoid robot.

With the help microcontroller (mcu) we have interact both the circuit. We used ATMEGA16 for cascading purpose. Output from the 'recognition' circuit is feed to microcontroller as an input and relative output is given to the 'synthesizing part.

E. g. as mentioned in the above task of 'toast', if we have to make it then...

We will save word 'toast' into digit '1' in the recognition part, means

If you speak 'toast' then output from 'recognition' circuit will be '1'.

Microcontroller will be programmed in such a way that if it gets input '1' at specific port then it have to send array of allophones to 'synthesizer' part. And of course that array will mean 'how light' as mentioned above.

Then synthesizer circuit will pronounce those allophones and it sound like 'how light?' time difference between the two allophones of single word may give you feeling of Question

Also we can interact some task with this, means if user say 'medium' then the task of lowering of 'bread tray' which were engaged to the heating coil will be scheduled in the microcontroller itself. Some ports of microcontroller may carry output to 'toaster' for control.

In the next article of utility we have given you some working examples of this **SPEECH INTERACTIVE SYSTEM** with some task to make this project more and more useful.

FUTURE CONSIDERATIONS

1. Speech and voice recognition security systems

If you want to use this speech interaction system in security system then you can also do this. You have to do only that just put 'face recognition system with this. 'Face recognition' system will provide security features and our 'speech interaction' will perform require task. Let's take one example for understanding this utility.

Suppose our task is to control 'robot' or just 'small toy car' but in security system, then we will allow user to used this system if passes through 'face recognition'. That is if user's images had been already saved in the database then only his login image may match with it. And if matched then user's login code will send to 'speech interaction system' from 'face recognition', and through that code specific commands which was allowed to that user will only work. Also his profile will be pronounced by 'speech interaction system' after login the user for better understanding.



This application is not very difficult because 'face recognition' is not new thing to understand. And hence we can consider this application.

2. Catching moving ball by blind person

Catching moving ball by blind person seems impossible, but it can be worked out with help of additional 'image processing' system with our 'speech interaction system'. 'image processing' system will guide the dynamic positions of ball and 'speech interaction' will order the path to blind person.

Image processing:

Through the camera it will detect the position changes of the ball, and robot will move accordingly for searching the ball with the help of image processing.

Algorithm is very clear; simply code will be programmed such that it will always search the same properties of the ball which had been previously appeared and will give changes in the positions of the ball accordingly to the robot. Robot will move accordingly with the help of interfacing between the robot & computer.

Actions to do in image processing:

1. Image acquisition
2. Image **Processing**
3. Data communication

Algorithm of Ball Follower:

Step 1:

Image acquisition, with the help of web cam mounted on the robot. through it image will be grab at specific time interval.

Step 2:

Find the center of the ball (ball is of definite color and that color will be different than background).

Step 3:

Locate the ball and assume its center as origin of new coordinate system which is impose on the image.

Step 4:

Now again & again check wheather ball center is moving coordianate system or not, by processing the input image. image will be taken at specific interval of time.

Step 5:

If ball position seems to be shift according to new coordinate system then move your robot accordingly.

Step 6:

Same procedure will be repeat again again.

Note:

It may be vary difficult to determine the dimenssions on the plane by which ball has been moved by image processing since its 3-D image processing & it require more than one camera in different directions.

On robot there will be pretty difficult to mount atleast two camera which may give us X- , Y- & Z- directional movements of ball on the plane.

But if camera is mounted on roof then it will become comparatively easier to give the exact position of the ball with respect to user.