Sri SivaSubramaniya Nadar College of Engineering
& Panimalar Engineering College
Affiliated to Anna University
TamilNadu India

VISION

Engineering solutions for the visually challenged

Team Details

Mentor

Dr V Seshadri Professor Emeritus *vision.csidc@gmail.com*
Sri SivaSubramaniya Nadar College of Engineering
India

Team Members

Arjun Vadhoola Seetharaman *vs.arjun@gmail.com*

Mayur Mudigonda *mupstein@yahoo.com*

Karthik Srinivasan *sambasivant@gmail.com*

Ashwin Kumar *kashwin@gmail.com*

1. Abstract

vision n.
1. the ability to see; inspired and idealistic ideas about the future.
2. a dream, an extraordinarily beautiful person
- The Oxford English Dictionary

Technological enhancements of our era have improved the quality of human life to an unprecedented extent. In this fast growing and ever shrinking global information scenario, we as technologists, researchers and computer scientists must devise ways of using these developments to benefit the society. We as technologists felt the need to visualize a Weltanschauung. There is an urgent and ever increasing need for sharing information and knowledge that is of prime importance aiming at societal transformation and empowering.

Immense work has lead to successful deployment of innumerable software and information systems that have changed the lifestyle of millions of people. Ubiquitous computing is the trend of the day. It is the duty of us to empower every human with necessary knowledge and keep him abreast with the happenings. The considerable advancement in technology, made a few academicians to start working towards improving Human Computer Interaction (HCI).

After a lot of brainstorming sessions and considerable literature survey, we formulated the essence of our project as *"VISION - Engineering solutions for the Visually Challenged"*. This is a conception of our ideology to give back something to humanity, especially to the visually challenged. Our visits to several Blind schools and Red Cross have motivated us furthers.

To the best of our knowledge our system is the first of its kind in trying to provide an integrated systems solution in India. We have tried to cater both to the knowledge enrichment aspects and also to alleviate some of the difficulties faced by the visually challenged.

2. SYSTEM OVERVIEW

2.1 Literature survey

The academicians in this community have for a long time felt the need of Computing for a social cause and societal transformation. With this intent various universities/companies set up the Human Computer Interaction (HCI) research labs that are ever burgeoning and meliorating towards achieving the goal of making computers help every human. Such research centers have, of late started exploring the possibility of how computers can be used effectively in interactive mode to cater to the needs of the physically disabled, especially the visually challenged. Researchers in this field, in collaboration with a number of blind/Braille organizations have developed a number of tools like refreshable Braille displays, text-to-speech synthesizers that have enhanced social and professional integration of the visually challenged. HCI community has in parallel evolved with state-of-the-art GUI that is packed with the "visual information metaphor" usable by the visually challenged. Researchers and industries, like IBM and others have developed specialized tools that cater to the visually challenged through multimodal and multimedia interfaces. Non-visual browsers, haptic interaction tools and sonification tools have also emerged. Such efforts are either content based or context based. However such tools suffer from proprietary issues and are quite expensive that very few visually challenged people can afford.

2.2 Overview

While working on this project we have tried to keep in mind the user's needs. Thus we have not limited our project by concentrating in building say one tool, a device or an application but have tried to keep the larger perspective of *going beyond the boundaries*.

Thus we ended up building a suite of applications to cater to the various needs, which we call '**VISION – An Engineering Solution'**. The suite consists of both applications of existing technologies and systems but also some indigenous tools developed through the course of the project.

What we intend to do through our project is to provide a suite of tools/applications that will either directly help the visually impaired in their regular life or indirectly in helping them learn and educate themselves. It also helps them to use modern technology and the power of the information age better.

We used an *Incremental Waterfall lifecycle* model while working on this project. The reason was that after each time we went through the lifecycle we kept finding newer ideas and better way to implement the older ones or the newer ones, which will become old soon. In certain modules we performed *eXtreme Programming (XP)* which suited the purpose.

The software design methodology we used in our design phase basically is UML. For the hardware aspects of the project we initially used scaled down prototypes.

At the beginning we drew *Use Case* diagrams, which played a very crucial role in our understanding of the system. The human interaction is very vital and crucial to the acceptance level and time. Hence to ensure the best possible design, each of us came up with our version of case diagrams for each interaction and then we chose the best features of the four.

The system we are developing is intended as an interactive tool for the visually challenged. Hence, it is necessary that all interactive components be modeled as objects. Thus all our analysis and design was object-oriented in nature. Since we are developing a large system, we made our design component based. Such an approach also facilitates easy packaging of various objects into components that are easily deployable, ready to use and easily upgradeable. The development is essentially based on the 'Incremental Waterfall Lifecycle Model'. For continuous improvement we included feedback mechanisms.

We sincerely believe that our efforts given a chance will prove to be really useful especially after the heartening results we got from our testing.

1. This system (package) is unique in the sense that it addresses a range of needs that haunt the visually challenged from reaching their true potential.
2. Our system is very cost efficient as the total cost if manufactured as individual pieces amounts to only 150$ which presumes interfacing with 10 devices/appliances.
3. We developed a simple but unique key skin called ABRAKS-Attachable Braille Key Skin, which maps the Braille –grade I onto the keyboard. Thus the user can use the regular keyboard by just placing this key skin, which we have made of a rubber polymer.
4. The learning system is a novel concept that uses the power of XML and Speech Recognition Software to provide learning aids and tutorials easily and effectively to the visually challenged.

3. IMPLEMENTATION AND ENGINEERING

The project can be divided approximately into four modules.

3.1 E-Learning System

The E-Learning System is one of the cores and fully tested module of the "VISION" suite that we have developed; it consists of a completely *networked learning system*, called 'Virtual Vision learning Center'. It can recognize discretely programmed commands and can provide study materials using a voice feedback system. It is targeted as a vital tool in educating the visually challenged through the voice enabled interfaces. It has a certain level of intelligence, the system is dynamic and new lessons can be integrated seamlessly.

It also helps in building an *Interactive Voice Response System*, thus making the computer more user friendly.

Working

The E-Learning system is a tool that can be used full fledged to educate the visually challenged in the subject of their interest. This networked system is essentially a client server architecture system. The client application runs on a low powered PC (essentially a thin client) that can recognize discretely programmed voice commands issued by the user. The user communicates with the system using a Bluetooth headset. A voice over will give the available list of subjects under the learning system; the user can then select a lesson by just saying the name of the lesson/topic.

Once a lesson/topic has been chosen, the user is allowed to choose sub topics and further down the hierarchy. These available topics under the lesson are listed using a voice over similar to the way in which the main lesson is initiated.

When the user has reached the desired sub section, the software reads aloud the contents (essentially the study material or information) of the section. The sections are stored in the form of MP3 files and are streamed from the server as and when it is needed. The user can also abort the system at any time and switch over to a new section desired. This tool also supports Question & Answer and Review Sections. A quiz material is also available with this system.

A major advantage of this tool is that it is user independent and can even recognize female voices, when it is actually programmed for a male. Another advantage is that, to use the system no training is required; the software guides the user completely.

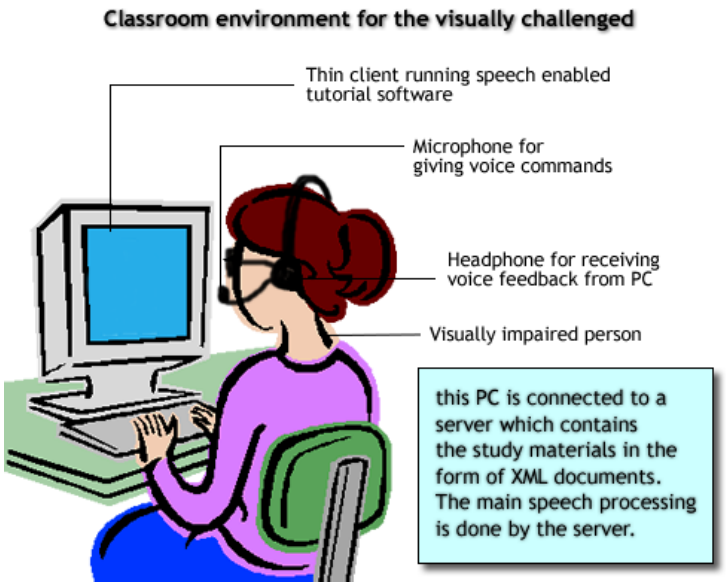


Figure 1. Classroom Environment.

Technical features

E-Learning system involves the design of a 'Client-Server' model as shown in Figure 2, in which the server stores the study material structured in the form of XML documents. The client performs the speech recognition part of the system from the user and gets data from the server as and when it is required/requested. XML parsing is done on the client to fetch the correct file. The server stores the hierarchy of XML files and the files are structured using the main file *root.xml.* This file along with the files associated with it can be changed or modified if the study materials are to be changed.

Sequential block diagram for E-Learning Application

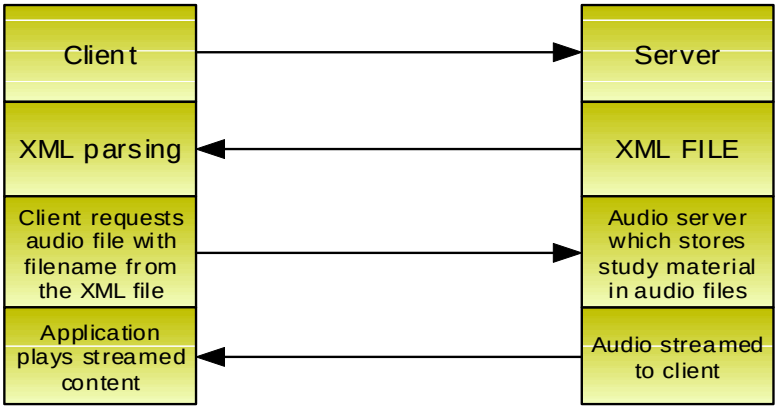


Figure 2. E- Learning Client Server Architecture

The final audio file, associated with a particular section of the tutorial containing the required information, is stored in the form of an MP3. The filename and location of this file is specified in the corresponding XML file. Thus a required MP3 file is opened and the audio is streamed to the client. We used XML documents in this system, as structuring of data is very simple unlike many conventional database systems. Moreover the XML documents can be easily restructured and modified to suit a particular schema. The XML parsing is quite simple and is done on the client side as and when data is needed. A voice feedback system, which essentially is a Text-To-Speech engine, reads out all the menu structure at each level and the selection is done using simple command recognition issued by the user.

The E-Learning System has been fully tested and is found to be exceptionally accurate. Accuracy is rated at around 93% and can be pushed to almost 97% when properly trained. The main reason for the accuracy to be less than 100% is because of the discrepancies associated in the transmission capabilities of the Blue tooth headset.

Another tradeoff that we had to compromise was that the Visually challenged, here in India found it difficult to understand the English in the voice feedback system, as it is essentially an American English TTS Engine.

This prompted us to evolve a Text-To-Speech Indian English engine that we are now working on. Sooner we would integrate our custom made TTS tool with the E-Learning system.

3.2 ‘ I SAY! ’ - Interactive Speech Enabled Home

For a visually impaired person, the ability to use his/her voice to operate devices is a great boon. Conventionally, the person is required to have a mental map of the room and physically operate specially made switches and buttons. With the option of giving voice commands to change the status of any device within a certain range would be a great boon for the visually challenged.

Hardware Requirements
This module has varying need of hardware components. They are as follows:

- RF Tags(Worn by the user)
- RF Receivers placed at suitable locations in the environment
- RF Transceiver attached to the computer
- RF Receivers attached to the devices to be controlled

RF-ID tag
As mentioned above the wearable RF-ID tag worn by every individual consists of a battery driven low powered RF module, a low power micro controller and a Dallas DS2411 Silicon Serial Number, which gives the tag a 64bit Unique ID

RF-ID Receiver
The RF ID receiver consists of a low power RF receiver tuned to the same frequency as the RF tag. The receiver is placed such that it will not be able to receive transmission from any other room. The RF-ID receiver maintains a data link with the main controller, which is the personal computer in this case. Different frequencies are used to prevent any contention between the RF tags and the computer-RF receiver link.

RF Communication
An interference free link is established between the computer and the device. Frequency hopping is used for eliminating error due to interference from other transmissions. The frequency of operation is 2.4 GHz, which is in the ISM license free band. A Laipac NRF2.4G wireless module is used for establishing data connectivity with the device. Data is transmitted in the form single byte information. FSK encoding is used for RF transmission. Interface with computer is through USB or serial interface.

Protocol Used

The RF-ID transmitter to receiver data link uses a 2400bps ASK (Amplitude Shift Keying) serial link operating at a frequency of 433.92MHz. The RF-ID receiver to the computer uses a 1200bps serial link with ASK encoding. The frequency of operation is 413 MHz, which removes problems of interference with the RF tags transmissions.

The system uses a multi-protocol wireless network set up in the house, which integrates seamlessly into the Virtual Vision Integrated Environment. The user (visually impaired/other) is uniquely identified using the wearable RF ID tags.

Contention Resolving

CSMA/CA is used as the contention-resolving scheme for multiple tags present in the same environment (room). Every tag before transmitting will check for any data transmission already in session. In such a case, the tag will choose a random back off period during which the unit will maintain radio silence and will try again after the back-off period. The back off period follows an exponential trend with a starting time of 224ms.

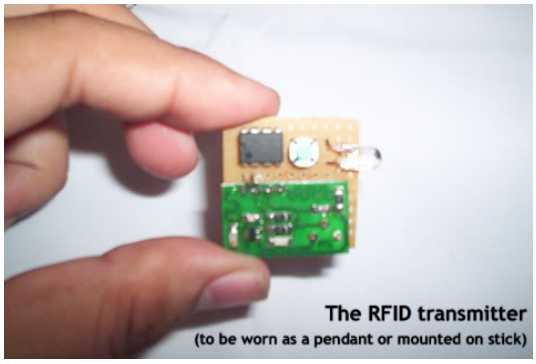


Figure 3. RFID Tx

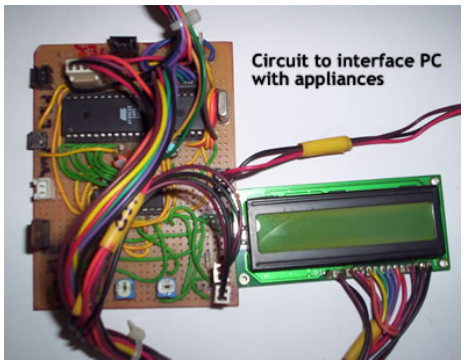


Figure 4. PC RF Tx

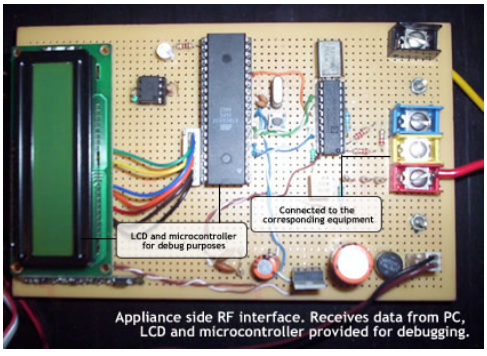


Figure 5. Appliance side Rx

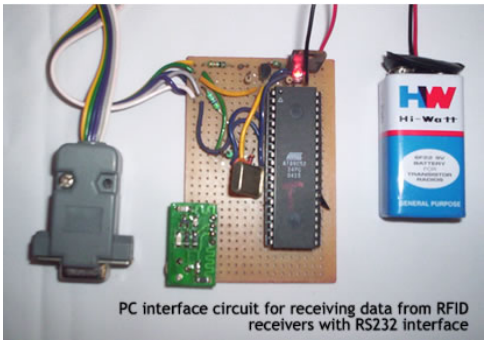


Figure 6. PC RF Rx

Working

Every user who interacts with the environment may be provided with a RF-ID Tag [Figure 3]. This tag is used to uniquely identify the user in the room as stated before. The monitoring software also gives a visual indication. The tagging establishes the spatial cue of the user in the environment.

The user's tag [Figure 3] transmits a message to the RF-ID receiver. The receiver [Figure 6] then transmits the UUID (Unique User ID) to the computer. At the computer a simulation [Figure 7] is runs which contains the environment's (house, office, etc) map. Based on this map the system will give directions or other data as requested by the user.

We have also incorporated a sub module which will generate optimal location points for the mutual exclusion of the receivers based on the blueprint/ map fed to the system. The location points are generated from the inputs using AI planning algorithms.

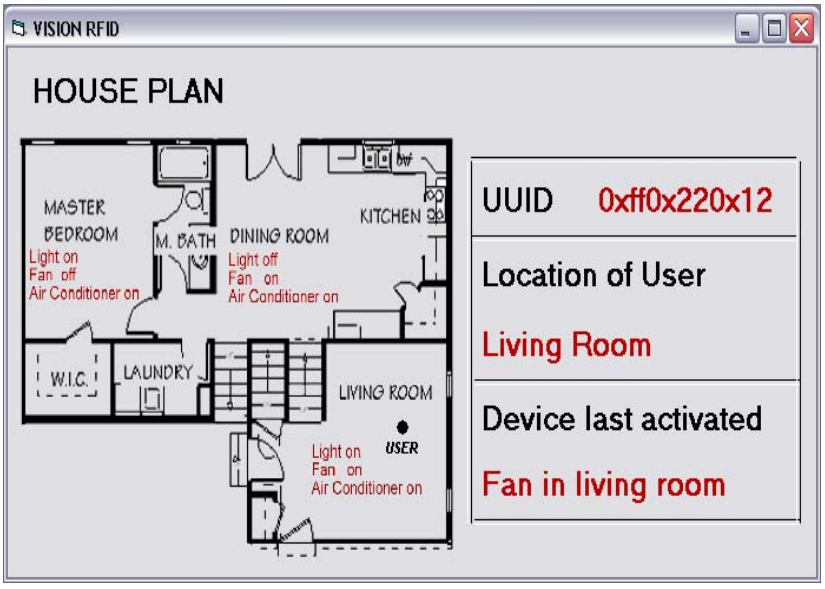


Figure 7. User Location Simulation

The user will also possess a Bluetooth headset, which provides an audio gateway service to the computer. The microphone of the headset is configured as the primary audio input device to the computer. A SAPI (Speech Application Programming Interface) based command recognition engine will run as a part of the *Virtual Vision Integrated Environment*. The engine is responsible for recognizing

any voice command given by the user. The software will automatically map the room in which the user is issuing the command from and the appliance to which he/she is referring.

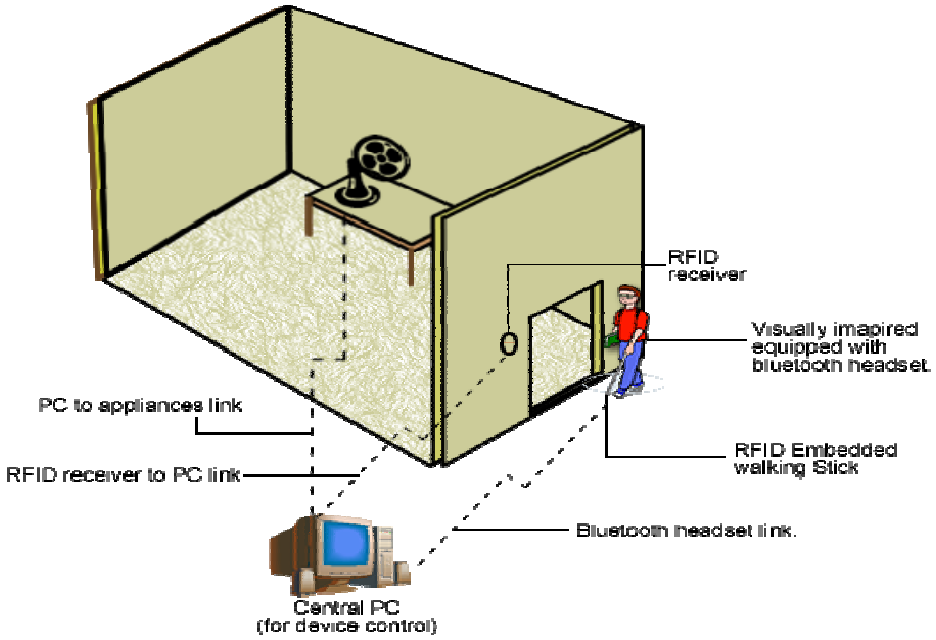


Figure 8. User triggering the receiver

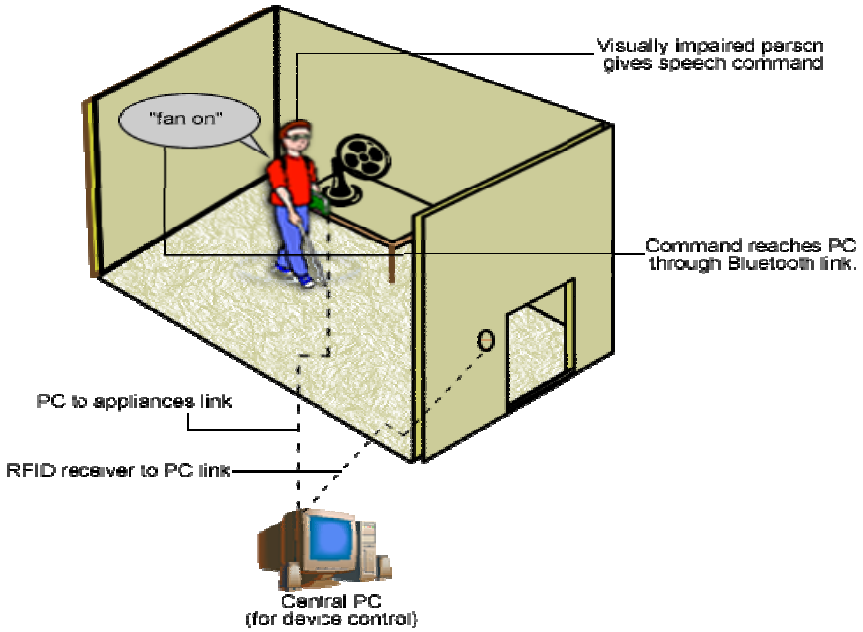


Figure 9. User issuing commands through the Bluetooth headset

The engine then calls a function that uses the SAPI and resolves the command issued by the user. It then looks up in the database and

sends out the required commands to perform these operations to the transmitter. The High power transmitter broadcasts the command to all the devices. But since the command consists of a unique id to identify each device only the required device will carry out the task specified through the command. Every device is also equipped with a low cost micro controller, which allows implementation of intelligent functions such as scheduled operation of the device, etc.

Power Line Communication- An Alternate Proposal

It is well known that the normal power cable has the capacity to transmit along with power at a maximum speed of 300 bps. There also exists the X10 protocol used for the same purpose. Thus while developing this module we made our protocols compatible with the X10.

Every device, which is to be controlled by the PC, is connected to a power line modem. The PC equipped with a low power line modem will be able to send commands to all the devices.

The power line communication protocol does not provide for addressing. All transmissions are sent as broadcasts. Hence addressing should be implemented in the higher levels. Every device is provided with a unique ID. More devices can be added by simply adding the device ID in the main controller program.

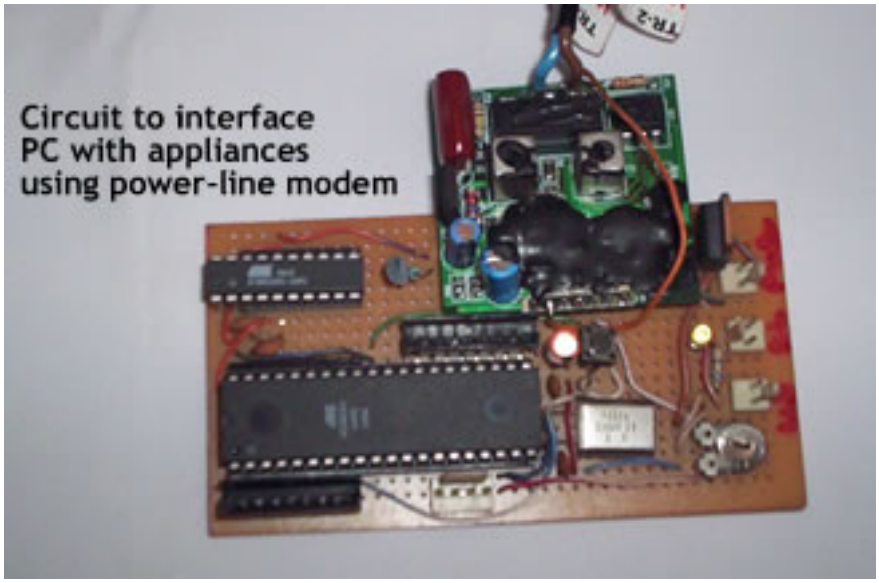
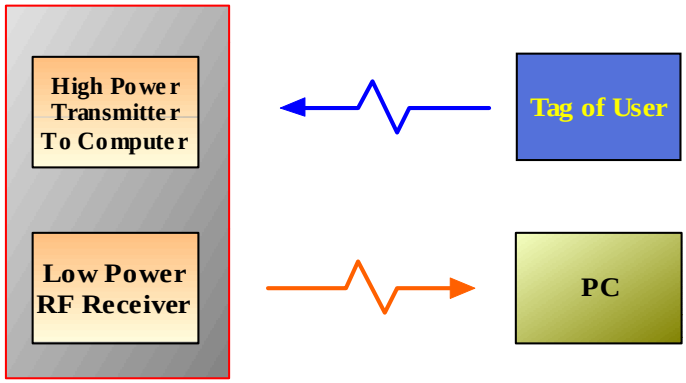
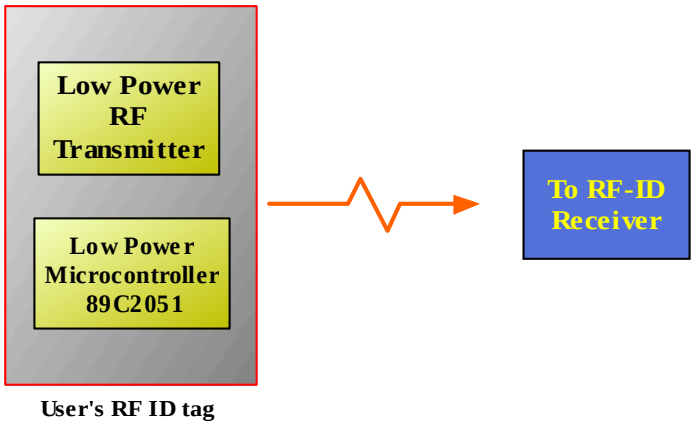


Figure 10. Power Line Modem

RF Receiver Placed Next to Entrance of A Room



RF ID Tag Worn by The Blind Person



A plausible scenario for serious system malfunction would be when the security system incorporated with the ISAY loses power. This raises the challenge of reconfirming the security settings. To compensate for such a situation, an auxiliary powered SOS unit is attached to the main computer and the security system that prompts the user to reconfirm the settings.

3.3 SODS – *SONAR Obstacle Detection System*

From our field study we found that a primary roadblock for visually impaired people is the detection and avoidance of collisions.

For this we developed the SONAR based Obstacle Detection System (SODS). This system consists of a detection unit that consists of ultra sonic transducers and the unit sends the ultra sound data to a Mobile Controller. The data is processed and a digital voice feedback is provided to the user over the ear phone.

The design of the range finder was done with the following requirements in mind.

- Rugged
- Should be small enough to be mounted on a cap
- Voice feedback

Working:

The range finder makes use of a PIC16f84 to do all the computational work involved. The range finder computes the distance of the nearest obstacle by calculating the time of flight of an ultrasonic burst transmitted by the device. The time of flight and the speed of sound in air, which is approximately 330 meters per second, is used to compute the distance of the possible obstacle.

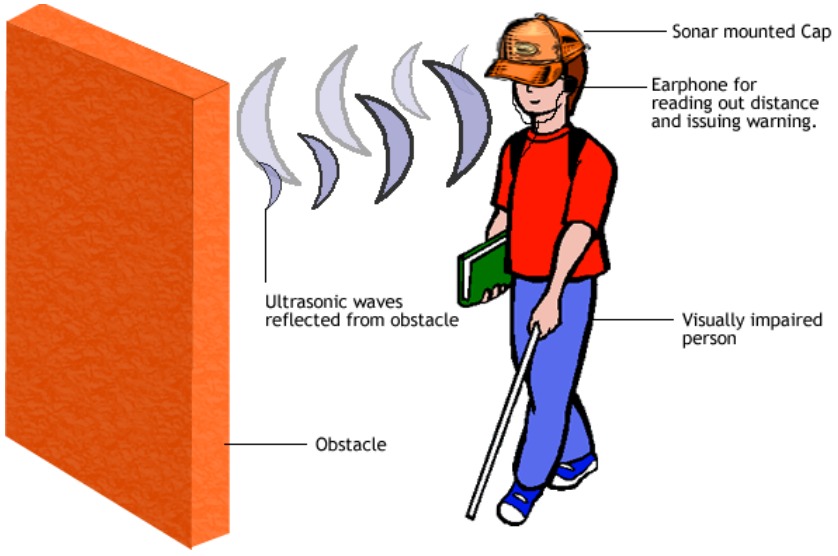
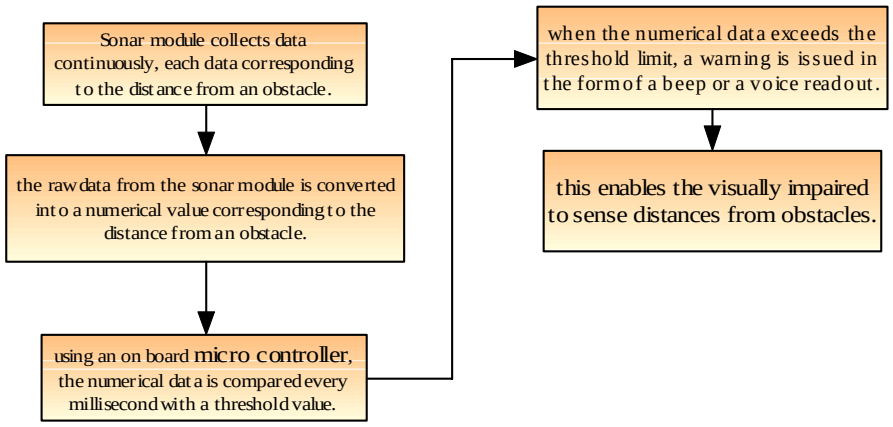


Figure 10. Ultrasonic obstacle detection.

The microcontroller sends seven consecutive pulses at a frequency of 25 kHz. The receiver waits till at least 4 of the transmitted pulses are received back. This threshold is set to prevent false triggering from spurious signals in the operating frequency.

Vocal output is given using *APR9600* voice recorder and playback IC. The IC will store all numerals between 0 and 9. Numbers are generated corresponding to the distances.
The controller *PIC16f84* is responsible for obtaining the signal from the ultrasonic transceiver, calculating the elapsed time period and the approximate distance of the obstacle from the person. The data is then readout through the earphone worn by the user.

Process flow for collision detection



BLOCK DIAGRAM FOR RANGE DETECTION

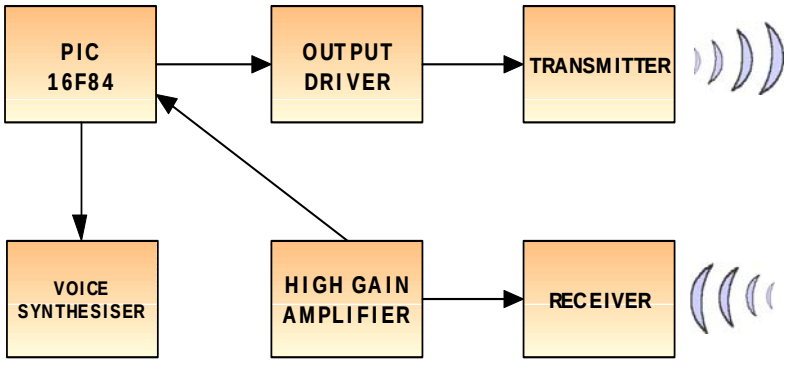


Figure 11. Obstacle Detection.

Cost and Accuracy:
The cost of the unit can be as low as 10$ in volume production. The unit has an approximate range of 5 meters and accuracy of 2 centimeters.

3.4 ABRAKS – Attachable BRaille Key Skin

The urge to improve the user-system interaction, motivated us to design a new Braille key skin for a conventional keyboard. The Key skin has been innovatively designed, at this stage to support the Braille Type/Grade1 keys. The Braille Type/Grade 1 skin essentially supports all Arabic numerals, English alphabets and few characters and does not support contractions (abbreviations) unlike Grade 2. We have designed the key skin quite intricately so that the mapping of the embossed keys on the key skin matches with those of the characters on the keyboard and are at the same time, quite comfortable for the visually challenged.

Our main intent in sticking with the Braille Grade 1 rather than Grade 2 was:

1) The visually challenged find it more complex to *only* read documents, manuscripts with Braille Type 1, as it consumes a lot of time to read simple words and sentences.
2) Had we designed a Braille 2 Type key skin, it tends to take away the privilege of the visually challenged to type/write what they intend to write as contractions do not account for all letters and words. Moreover the language fluency of the individual is deprived using a Braille 2 skin.

The Design of the Key skin is ready and we are now in the process of manufacturing them. We plan to either make the embossed key skin using high quality rubber polymer. The key skin can be manufactured at around 5$ if mass produced.

However, we believe that only after the visually challenged people's experience is assimilated, there is any chance of providing better mapping and spatial cue than the one that we have conceptualized.

3.5 Testing And Validation

The testing process for the project has been a continuous process. The first module to be completed and alpha tested was the e-learning system. This module was designed and tested at the labs of

IIT Mumbai for a contest called 'Drishti 2005'. The inputs given by the judges were analyzed and incorporated in this version.

The 'I SAY!' module was tested incrementally; first we developed a simple five piece module interfacing a single device which worked out well. There were many issues regarding interference when placing more than one receiver in a mutually overlapping position which we resolved through higher end addressing and random exponential back off while transmitting from the RF-ID tag. We are currently testing the module with the X10 compatible protocol using a PLM.

The ABRAKS module has been a very challenging module for us. We are experimenting currently with various materials for the key skin. We are trying to test the durability and the ergonomics of the ABRAKS.

4. Summary

"There is always scope for improvement". This has been our motto all through out, this being the case we would like to briefly state our views with respect to the future work on this project.

Our project currently stands at the threshold of completion. We have to complete our beta testing processes for the device control. The e-learning system is being optimized for efficiency. The device control is being interfaced with more than five devices and also trying to resolve interferences.

First and foremost in all our testing (In India) the acceptance level was reduced significantly because of the lack of a customized speech engine i.e. an Indian Language Speech Engine. We have been spending considerable time on this and have felt that there is a significant requirement towards developing a speech engine and hope to develop software to design speech engines so as to make our project *language independent* quite like Java one may say.

Also in the pipeline is Braille to Speech converter. We feel that if such a package could be developed then it would definitely increase productivity and ease of using computer systems.

We plan to create an E-Learning tutorial for Encyclopedias in a similar fashion.
We also want to enhance the quality of service of such streaming audio from remote locations and capabilities to learn lessons/tutorials from the Web.

5. References

Through the course of our project, we have referred a lot of manuals, whitepapers and collaborated with the academia.

[1] SAPI- Microsoft Developer Network 2005, http://www.msdn.com

[2] TTS Whitepaper - http://www.sakramentspeech.com/products/tts/

[3] Indian TTS Whitepaper – http://www.blissit.org

[4] Datasheets and reference designs for AVR, APR9600, PIC16f84.

[5] Software Engineering – A Practitioners Approach, Roger S Pressman, 5th Edition, McGraw Hill.

[6] UML Distilled: A Brief Guide to the Standard Object Modeling Language, Third Edition, Addison Wesley.

[7] Dr C Aravindan, Head of Department, Department of Computer Science, Sri SivaSubramaniya Nadar College of Engineering.