

Communication Between Visually Impaired Individuals

¹ Vishakha Hiremath, ² Vineeta Gejji,

¹PG Student, ²Professor,

¹Department of Electronics and Communication Engineering,

¹Gogte Institute of Technology, Belagavi, India.

Abstract— One of the most important and challenging issue in human interaction is the communication between the visually impaired individuals. The method of communication is totally different from the language of other non-disabled people. Few cases (i) A blind person speak but the deaf cannot listen; (ii) Deaf person makes sign language signs but the blind person cannot see it. Thus this paper study on the issues and problems that are faced by the visually impaired people and finds a solution by making use of JSAPI module and Sphinx module. The proposed system offers solution to the challenges.

IndexTerms— JSAPI,Sphinx,Speech Recognition,Speech synthesis.

1. Introduction

1.1 Overview

According to the World Health Reports (WHR), there are millions of blind and deaf impaired individuals who are suffering from lack of improper communication. Blindness, deafness and dumbness may be traumatic which can later leads to some depression and frustration. Absence of attention to the visually impaired/hard of hearing needs could prompt segregation in a society community. It may lead to many talented visually impaired people out of the profitable individuals from the society. And it may also affect the educational performance of the children. Besides, the children with these kind of disabilities in the developing nations are hardly to complete their education. Specifically, elders with visually impaired face lack of job opportunities in day today life. The rate of blind people who are unemployed is (58%) and deaf is (18%). Educational things which can later translate to employment opportunities and become independent are very rare for the blind and deaf impaired individuals. In addition, the interaction communication between these two classes of debilitated individuals is extremely difficult and in some cases it is highly impossible. It is unfortunate fact that having good devices and IT synergistic systems that can be used for these people to overcome the issues or the problems that are faced to seek their education and improve there communication opportunities which will allow them wind up as a profitable members of the society. Thus, this paper allows the blind or the disabled people come into light and make there communication stronger and finds an chance that will interface with the distinctive modalities which are taken from the same module or the domain which will be understandable by both the communities that is the blind and also from the deaf individuals. The solution is done by making use of the Java wherein TTS (text-to-speech) and STT (speech-to-text) conversions are carried out in different sets of modules.

1.2 Objectives

- Blind person can send mail to any other person just using his speech.
- Blind person can chat with the other person easily where the communication data will automatically converted to STT.
- Deaf/Dumb person text message will be automatically converted to a speech for blind person etc.

The analysis of the role of a proposed system and the identification of the requirements that it should meet is known as system analysis. SA is the starting point for system design. This term is most consistently used as a piece of the setting of business programming, where programming engineers are habitually classed as either systems analysts or programming engineers. The frameworks investigators are in

charge of recognizing necessities (i.e. frameworks investigation) and delivering an outline. The software engineers are then in charge of executing it.

1.3 Drawbacks in the existing system

- The blind people should have bellyscript knowledge to operate the bellyscript keyboard.
- The blind people need someones help to operate the system.
- Deaf people cannot interpret voice messages.

2. Related Work

Individuals are the social creatures. While the individuals with visual and hearing problem can figure out how to impact their thoughts on others and share their views. The low power consumption and small size ARM based computer for example, Raspberry Pi has opened the gateways for some of the embedded systems including a heart watching method and an assistive course structure for the outwardly impeded. This paper [1] shows the outline, model and testing of a minimal comfort and speaker device with a braille refreshable show for the correspondence between two people (either being nearly deaf outwardly weakened) that has both, an also ease, and various potential results for advance change on the ARM-based PC system.

In reality, Information Technology with its frontline procedures, for instances, for flexible and distributed computing has an enhancing part on the intercommunication among the population with the different inabilities and ordinary students from one side and among the handicapped people themselves who have the same and distorted impairments. It is proposed [2] by a board structure upon three primary current progressions: phones, cloud resources and informal organization to give a predictable communication between the blind and deaf impaired individuals.

The paper [3] undergoes for theoretical examination and also with aim of remote correspondence future that fulfils the outwardly disabled and partially deaf people. The contraption got a joined DTMF gadget with large number of sensors like Heartbeat, Vibration, Raindrop and Ultrasonic sensor etc. Here the whole system is planned with estimated catches called Sensors which is to be utilized by the user. The short working can be clarified as; at whatever point a man is holding stick has high pulse rate or low heartbeat rate or is in any horrifying condition a message will be sent to the concerned individual and the GPS appended to the hindered will get initiated.

The fundamental aim of the paper [4] is to manufacture a system which helps to make a life of the blind and deaf individuals. This system consolidates two modules: one is giving direction and

another is obstacle recognizable proof. The framework indicates course by means of PDA utilizing voice charge also, to recognize any impediment in transit, ultrasonic sensor is utilized. In a word, it is a system which gives an aggregate lead to help an apparently hindered individual accomplishing any objective safely.

The paper [5] shows the analysing and the demonstration and investigation besides the working of the microcontroller it is being tested which depends on the smart assistive framework which can be easily used by the impaired people. This device includes the haptic and audio feedback from which the user can make selections. The gadget is also small and main goal is to alert the customer when objects are accessible. Partition estimations, between the client and conceivable obstructions, are performed utilizing ultrasonic echolocation and the data is given by its sensor and is set up by a microcontroller, which additionally handles the feedback part.

Conveyability of outwardly impaired individuals is restricted by inside their environment. Moving securely and unhesitatingly in a metropolitan zone with no human help is a monotonous work for vision misfortune individuals. This paper [6] proposes hypothetical show and a framework idea to give an electronic guide to visually weakened individuals. This work depend on building up a contraction and a Bluetooth headset (wearable), for daze individuals that causes them to discover their way in this world.

The headset and the proposed stick are combined utilizing Bluetooth Stick. In this manner IOT help blind individuals to discover their way back home by the correspondence between the walking stick and the headset, which are Bluetooth, matched and empower the communication between gadgets.

Visually impaired kids and grown-ups experience issues in taking part in positive social communications. This investigation surveys social fitness in located and outwardly debilitated individuals and to propose a novel interventional system in outwardly disabled youngsters.

This paper [7] is composed that surveys the capacity to start and support cooperation with the experimenter while performing free hand developments. Two members and experimenter kinematic information were recorded with a movement catch framework.

Disabled people exhibit a poorer correspondence stream with the experimenter than located individuals, which shows a less proficient social communication. The measure of correspondence between the two agents brought about a critical change after the interventional program.

Visually debilitated individuals discover awesome trouble in moving around openly without a human guide, particularly in another landscape. iWalk is a wise strolling stick exceptionally intended for outwardly incapacitated people to enhance and simplicity route.

iWalk utilizes ultrasonic and water sensors to distinguish hindrances and water ahead. The indicator incorporated in that actuates a particular bell on the off chance that it distinguishes water. Those system accompanies a remote RF remote control that extraordinarily sounds buzzer when squeezed, that aides in finding the stick in the event that it gets lost. The effectiveness of the proposed arrangement has been satisfactorily tried on a dataset made out of randomized areas with hindrances and water. The paper [8] proposed algorithm accomplishes a general effectiveness of 90% recognition rate for water and ultrasonic sensor and 85.75% for the RF remote control.

3. Design

It is flow by which a computer distinguishes spoken words. Speech recognition (SR) frameworks enable individuals to control a computer by addressing it through a microphone, either entering

content, or issuing orders to the computer, e.g. to stack a specific program, or to print a record.

Speech recognition enables you to furnish contribution to an application with your voice. Much the same as clicking with your mouse, composing on your console, or squeezing a key on the telephone keypad gives contribution to an application, Speech recognition enables you to give enter by talking. In the work area world, you require a receiver to have the capacity.

Speech is a characteristic method of communication for individuals. Most important behaviour and amid early youth, without guideline, and we keep on relying on discourse correspondence for the duration of our lives. It comes so normally to us that we don't understand how complex a system behaves whose activity isn't simply under cognizant control yet in addition influenced by factors going from childhood to enthusiastic state.

Accordingly, vocalizations can change broadly as far as their complement, elocution, explanation, unpleasantness, nasality, pitch, volume, and speed; in addition, amid transmission, our sporadic discourse examples can be additionally misshaped by foundation clamor and echoes, and also electrical qualities (if phones or other electronic hardware are utilized). Every one of these wellsprings of changeability make speech recognition, considerably more than speech generation, an exceptionally complex issue.

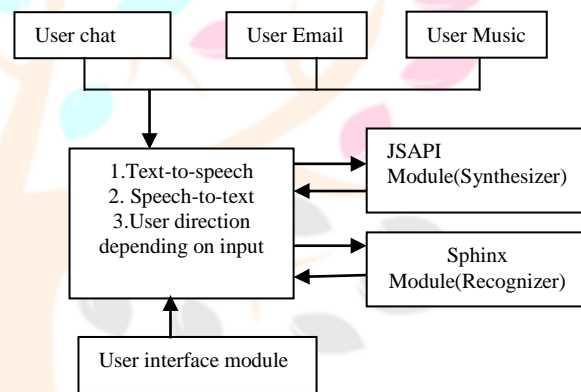


Fig 1: System design

3.1 System Overview

Discourse acknowledgment structures can be disengaged in a couple of one of a kind classes by depicting what sorts of articulations they can see. That gathering relies upon the way that one of the difficulties of ASR is the ability to choose when a speaker starts and finishes an enunciation.

1. Segregated Words

These recognizers as a rule want every expression to have calm (absence of a sound flag) on BOTH sides of the example window. Regularly, these frameworks have "Listen/Not-Listen" states, where they require the speaker to hold up between articulations (for the most part doing preparing amid the stops). Detached Utterance may be a superior name for this class.

2. Connected Words

Interface word systems (or all the more effectively 'associated articulations') are like Isolated words, however enable separate expressions to be 'run-together' with an insignificant interruption between them.

3. Continuous Speech

Ceaseless acknowledgment is the accompanying stage. Recognizers with ceaseless discourse limits are likely the most difficult to make since they should utilize unprecedented methods to choose enunciation limits. Consistent discourse recognizers empower customers to talk ordinarily, while the PC chooses the substance. On a very basic level, it's PC correspondence.

4. Spontaneous Speech

There seems, by all accounts, to be an assortment of definitions for what Spontaneous speech really is.

5. Voice Verification/Identification

Some ASR frameworks can distinguish particular clients. This report doesn't cover check or security frameworks.

A talk synthesizer changes over created content into talked lingo. Talk amalgamation is additionally suggested as substance to-talk (TTS) change. The significant strides from content are as per the going with:

Structure examination: process the data substance to comprehend where fragments, sentences and differing structures begin and end. For most tongues, feature and organizing information are utilized as a bit of this stage.

Content pre-planning: investigate the information content for momentous makes of the lingo. In English, outstanding treatment is required for shortened structures, acronyms, dates, times, numbers, cash aggregates, email addresses and different particular structures. Unmistakable languages require intriguing getting ready for these structures and most vernaculars have other specific necessities.

Remaining methods that change over the talked content to speech are :

- **Text-to-phoneme transformation:** change over each word to phonemes. Different lingos have unmistakable courses of action of sounds (various phonemes). Such instance, Japanese has less phonemes fusing sounds not found in English.
- **The attempt for sound pattern analysis:** It process the sentence structure, words and phonemes to choose appropriate prosody for the sentence. Prosody consolidates a noteworthy number of the features of talk other than the indications of the words being talked. This joins the pitch (or melody), the arranging (or beat), the halting, the talking rate, the complement on words and various distinctive features. Amend prosody is basic for impacting talk to sound right and for precisely passing on the significance of a sentence.
- **Waveform production** is the sound unit and pattern data are utilized to deliver the sound waveform for each sentence. There are numerous manners by which the discourse can be created from the phoneme and prosody data. Most present frameworks do it in one of two routes: link of pieces of recorded human speech, or formant blend utilizing signal handling procedures in light of information of how phonemes sound and how prosody influences those phonemes. The subtle elements of waveform age are not ordinarily critical to application engineers.

3.2 Methodology

3.2.1 Working of Speech Recognition

Since it has examined a portion of the essential terms and ideas associated with speech recognition, we should assemble them and investigate the discourse recognition process functions.

A discourse acknowledgement comprises of various parts. These are gained from information, utilizing a Speech Corpus comprising of accounts of speech and their literary translations. It figures out how to make correspondences amongst sounds and words.

Signal Processing

This methodology the signs recorded by the speaker into Feature Vectors that give a see of what is occurring in the discourse flag, focusing on features that are critical to discourse acknowledgement. Usually, 100 component vectors for each second are conveyed.

As you can in all likelihood imagine, the discourse acknowledgement motor has a to some degree complex endeavor to manage, that of

taking rough solid data and making a translation of it to apparent substance that an application gets it.

The huge parts we need to talk about are:

- voice input
- Grammar
- Acoustic Model
- acknowledgement content

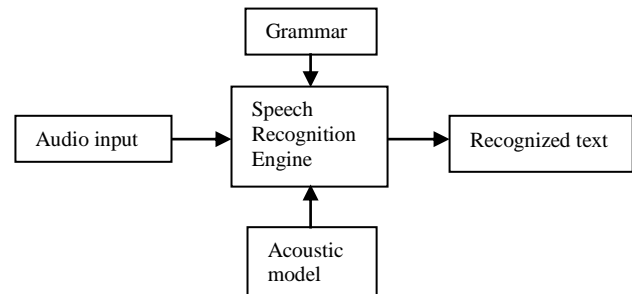


Fig 2: Recognition module

3.3 Speech Synthesis Limitations

Speech synthesizers can make blunders in any of the preparing steps depicted previously. Human ears are very much tuned to recognizing these blunders, so watchful work by engineers can limit mistakes and enhance the discourse yield quality.

This numerous approach for JSAPI and the Java Speech Markup Language (JSML) to get an application designer to enhance the yield nature of a discourse synthesizer.

The JSML characterizes how to markup content contribution to a discourse synthesizer with data that empowers the synthesizer to improve the discourse yield quality. In a portion of its highlights which upgrade quality include

- Ability to stamp the begin and end of sections and sentences.
- Ability to determine articulations for any word, acronym, shortening or other exceptional content portrayal.
- Explicit control of stops, limits, accentuation, pitch, talking rate and din to enhance the yield prosody.

These highlights enable a designer or client to supersede the conduct of a discourse synthesizer to revise the greater part of the potential mistakes portrayed previously.

4. Implementation details

4.1 Module description

4.1.1 Sphinx4

The Sphinx-4 system has been composed with a high level of adaptability and particularity and speaks to a module that can be effectively supplanted, enabling scientists to try different things with various module usage without expecting to adjust different bits of the framework.

Whenever before or amid the acknowledgment procedure, the application can issue Controls to every one of the modules, adequately turning into an accomplice in the acknowledgment process. The Sphinx-4 framework resembles most discourse acknowledgement frameworks in that it has an expansive number of configurable parameters, such as inquiry pillar estimate, for tuning the framework execution. The Sphinx-4 ConfigurationManager is utilized to design such parameters. Not at all like different frameworks, in any case, the ConfigurationManager additionally enables Sphinx-4 to progressively stack and design modules at run time, yielding an adaptable and pluggable framework. Utilizing the ConfigurationManager, however, it is conceivable to reconfigure Sphinx-4 to develop an alternate FrontEnd that produces Perceptual Linear Prediction coefficients (PLP) without expecting to change any source

code or to recompile the framework. To enable applications and designers to track decoder measurements, for example, word blunder rate, runtime speed, and memory utilization, Sphinx-4 gives various Tools. The tools which are exceedingly configurable, enabling clients to play out an extensive variety of framework investigation. Besides, the Tools additionally gives an intuitive runtime condition that enables clients to change the parameters of the framework while the framework is running, considering fast experimentation with different parameters settings.

Sphinx-4 additionally offers help for Utilities that help application-level handling of acknowledgment comes about. For instance, these utilities incorporate help for acquiring result grids, certainty scores, and normal dialect understanding.

4.1.2 FrontEnd

The inspiration driving the Front End is to parameterize an Input signal (e.g., sound) into a progression of yield Features. This incorporates no less than one parallel chains of replaceable passing on flag getting ready modules called Data Processors. Supporting various chains licenses simultaneous computation of different sorts of parameters from the same or assorted information signals. This engages the making of structures that can in the meantime decipher using unmistakable parameter forms, for instance, MFCC and PLP, and much parameter composes got from non-talk banners, for instance, video.

4.1.3 Dictionary

The Dictionary gives pronunciations to words found in the LanguageModel. The elocutions break words into successions of sub-word units found in the AcousticModel. The Dictionary interface additionally underpins the characterization of words and takes into consideration a solitary word to be in different classes.

The different executions advance for use designs in view of the measure of the dynamic vocabulary.

For instance, one execution will stack the whole vocabulary at framework introduction time, while another usage will just acquire elocutions on request.

4.1.4 Acoustic Model

The Acoustic Model module gives a mapping between a unit of discourse and a HMM that can be scored against approaching highlights gave by the FrontEnd. Likewise with different frameworks, the mapping may likewise consider relevant and word position data. For instance, on account of triphones, the setting speaks to the single phonemes to one side and right of the given phoneme, and the word position speaks to whether the triphone is toward the starting, center, or end of a word (or is a word itself) into consideration the meaning of Acoustic Models that contain allophones and in addition Acoustic Models whose settings don't should be contiguous the unit.

4.2 H/W configuration

Processor	Intel
Floppy Drive	1.44 MB
Key Board	Standard Windows keyboard
Mouse	Two or Three Button Mouse
Hard Disk	40GB
RAM	2GB
Speed	2.20Ghz

4.3 S/W configuration

Operating System	Windows 7/2009/Xp
Front End	Java swings
Scripts	Java Script

Software used

JCreator IDE

5. Snapshots of Results

In the synthesizer engine, SynthesizerModeDesc command is used to createSynthesizer technique. Later Allocate and Resume is done the program and speakPlainText is Generated and then Dellocate technique liberates the synthesizer assets.

Snapshot 1: Music Section output for the blind user



Figure 1: Music section output

Snapshot 2: Email Section output for the blind user



Figure 2: Email Section output for the blind user

Snapshot3: Blind user chat section .

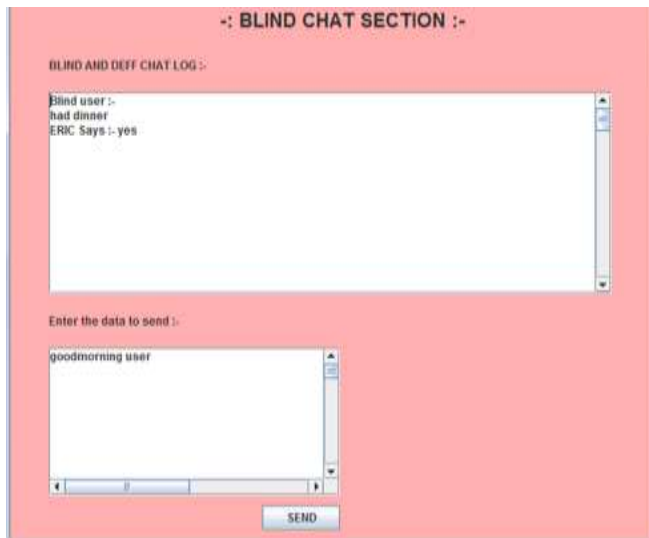


Figure 3: Blind chat section output

Snapshot4: Deaf user chat section.

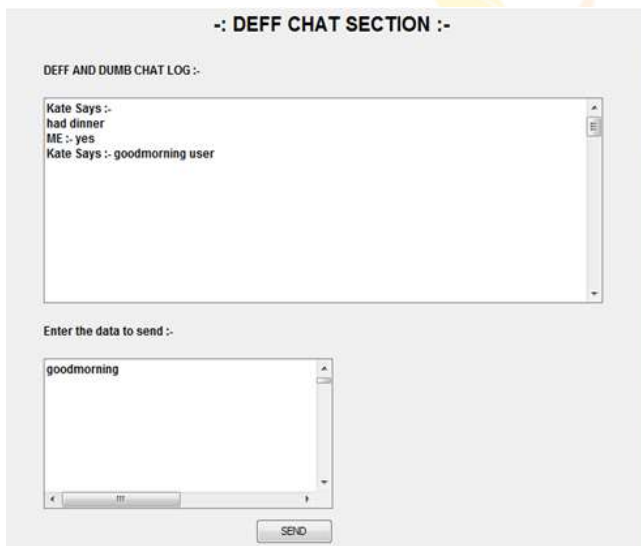


Figure 4: Deaf chat section output

Results are the articles that the Recognizer returns when speech is distinguished. It analyzes speech by speculating on plausible counterparts for what a client has said. The Result question really contains the greater part of the "pursuit ways" that the Recognizer has crossed for a given square of discourse. It contains ways that have come to their "last state" (implying that it's most likely toward the finish of a sentence or long interruption) and in addition "dynamic ways" that haven't yet achieved last state. Essentially, the Result question is an accumulation of scored surmises that the PC makes about what has been stated, and the protest has techniques for your application to mine through this gathering in various ways. The `getBestFinalResultNoFiller()` is the strategy utilized most in the demos, and it is utilized to maintain a strategic distance from any fractional sentences in the content yield. Fundamentally, the program holds up until the point that it is sure of a completed expression before it hands off a literary figure. Another conceivable technique, `getBestResultNoFiller()` appears more lenient (yet maybe less exact) in that it endeavors to restore the most astounding scored result that has achieved a last state, however in the event that it doesn't locate a best last outcome, it is cheerful to restore the dynamic outcome with the most astounding score. There are numerous different

strategies for controlling the Result returned by the Recognizer, including approaches to delve into the hunt ways, words, and scores of Tokens to discover the N-best results.

6. Conclusion

This project analyses the system for the interaction between visually impaired individuals and their compelling cooperation with the computer. The system consolidates a module which is equipped for perceiving user motions and makes an interpretation of them utilizing content to-voice and voice-to-content applications.

7. Future Work

Future work will focus around the extension of the created modules with a specific end goal to help bigger vocabularies and empower more common communication of the users.

Moreover, the structure of the utilized modalities ought to be contemplated more to uncover their between conditions and adventure their reciprocal nature all the more successfully.

References

- [1] Ramirez-Garibay, Fernando, et al. "MyVox—Device for the communication between people: blind, deaf, deaf-blind and unimpaired." *Global Humanitarian Technology Conference (GHTC), 2014 IEEE*. IEEE, 2014.
- [2] El-Gayyar, Mahmoud, et al. "Social network framework for deaf and blind people based on cloud computing." *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*. IEEE, 2013.
- [3] Shrivastava, Prasun, et al. "Medico stick: An ease to blind & deaf." *Electronics and Communication Systems (ICECS), 2015 2nd International Conference on*. IEEE, 2015.
- [4] Uddin, Md Ashraf, and Ashraf Huq Suny. "Shortest path finding and obstacle detection for visually impaired people using smart phone." *Electrical Engineering and Information Communication Technology (ICEEICT), 2015 International Conference on*. IEEE, 2015.
- [5] Munteanu, D., and R. Ionel. "Voice-controlled smart assistive device for visually impaired individuals." *Electronics and Telecommunications (ISETC), 2016 12th IEEE International Symposium on*. IEEE, 2016.
- [6] Mala, N. Sathya, S. Sushmi Thushara, and Sankari Subbiah. "Navigation gadget for visually impaired based on IoT." *Computing and Communications Technologies (ICCCT), 2017 2nd International Conference on*. IEEE, 2017.
- [7] Cappagli, Giulia, et al. "Assessing social competence in visually impaired people and proposing an interventional program in visually impaired children." *IEEE Transactions on Cognitive and Developmental Systems* 20 (2018): 1-1.
- [8] Olanrewaju, Rashidah Funke, Muhammad Luqman Azzaki Mohd Radzi, and Mariam Rehab. "iWalk: Intelligent walking stick for visually impaired subjects." *Smart Instrumentation, Measurement and Application (ICSIMA), 2017 IEEE 4th International Conference on*. IEEE, 2017.