International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-4 Number-1 Issue-14 March-2014

Hands Free Browser-'An Interactive Speech Browser for Visually Handicapped'

Sachin Kapoor[1], Kanika Mittal[2], Adit Gupta[3]

Abstract

These days the technology of browsing through voice is increasing rapidly. The paper concentrate on voice based approach of browsing rather than operating on textual mode, which uses speech recognition and speech synthesis methods or algorithms to implement speech technology. Now-a-days the learning over the web has become an important mode of education. However, the persons with no vision i.e. blind persons who represent a major population can't access the content over the web efficiently in the current browsers, as they are unable to see the different links which are present in the web pages. This paper propose a system which allows these people to access the web, it comprises of various modules, such as: Speech Recognition, Speech Synthesis, Text to Braille and Braille to Text conversion. The paper also summarizes the algorithms used for above modules.

Keywords

Speech Recognition, Braille, Speech Synthesis, CELP, HMM, DTW, MBROLA.

1. Introduction

Web surfing has now become an integral part of a large population across the world due to the never ending advancements in the internet. It affects the livelihood of the people to a greater extent as well. Since the internet is an important system for information retrieval, it is still not in reach to people with no vision, because most of the web content is present visually therefore this becomes an obvious source of difficulty for them. More importantly, overall content of the web page is not completely represented by digital contents [1, 2]. The major difficulty for visually impaired is that they are unable to see the positions of their mouse on computer screen i.e. mouse pointers or text cursors.

Manuscript received January 06, 2014.
Sachin Kapoor, Bhagwan Parshuram Institute of Technology.
Adit Gupta, Bhagwan Parshuram Institute of Technology.
Kanika Mittal, Bhagwan Parshuram Institute of Technology.

Without the use of proper and assistive technologies, surfing through hyperlinks along with locating different elements in web forms such as textboxes, radio buttons seem almost impossible. Data rate through Speech communication is higher therefore this technology is preferred over the other [2]. Speech technology made it easy for a blind person to communicate over the web and also utilizing the other benefits fruitfully.

A teacher uses conventional methods of braille dots to teach visually impaired people therefore there is no other method for these people to gain knowledge [3]. They cannot see the screens and links present on the web pages in the current browsers neither they can enjoy web services over the Internet like sending and receiving e-mails. Visually impaired people are not able to control the functionalities of the browser. It is very complex task for these people to physically learn all the functionality of the web browser [3]. Voice recognition helps the visually impaired learners to access multimedia content more effectively.

A voice-enabled web browser is a web browser program that can be operated using voice commands, in addition to normal keyboard and mouse operations. A webpage-reader program is integrated, so that users can surf the web based purely on voice. It uses a Text-to-Speech (TTS) module or commonly called as Speech Synthesis technology, which produce speech sounds from corresponding texts. The webpage reader programs aims at reading the text more accurately and precisely rather than naturally. Automatic Speech Recognition (ASR) or simply speech recognition software allows mapping from sounds to texts. ASR provides an alternative means of communication through voice browser where control and input methods uses speech technology [2, 3]. The browser also incorporates BTT and TTB conversion as braille characters are elementary part of impaired person as they learn reading and learning via this means only. Therefore this paper made it easy for these people along with their teachers to interact with each other online. The other user over the web can also interact with blind people via this

speech technology. The reading, writing, and surfing over the net will be easy from the blind perspective.

2. Literature Review

Worldwide Web was the first browser invented in 1990 by Tim Berners-Lee, later it was renamed as Nexus. Marc Andreesen, three layers later i.e. in 1993 developed a browser which sparked the internet boom of the 1990s. It was easy to use and install with the release of Mosaic. Mosaic was later name as Netscape. The Midas WWW browser Based on Motif/X was released in 1992 by Tony Johnson. It allowed viewing of PostScript files and even handled compressed PostScript. The first web browser which worked on Windows 3.1, NT 3.5, and OS/2 called Cello was released on 8 June 1993; it was developed by Thomas R. Bruce of the Legal Information Institute at Cornell Law School. In 1995 .Udi WWW was released in 1995, it was the first web browser that was able to handle all math tags along with HTML 3 features. Its further version 1.2 was released in April 1996. Microsoft, which had thus far not marketed a browser, finally entered the fray with its Internet Explorer product (version 1.0 was released 16 August 1995), purchased from Spyglass, Inc. This began what is known as the "browser wars" in which Microsoft and Netscape competed for the Web browser market. Since the web is an ideal source of information, it is still far from the people with low vision. So development took place in incorporating speech technology in browser, which will allow the blind persons to interact with the web through speech commands. M. Edgington and A. Lowry presented a set of novel algorithms for the signal modification component of concatenated text-to-speech systems [4]. An Optical Braille Recognition (OBR) system consists of basic module: dot localization and segmentation, and finally dot recognition and conversion. Among issues that must be taken into consideration when implementing an OBR system are factors that negatively influence the identification process, such as lighting conditions, page placement in the scanner, and page movement [5].

3. Proposed Technology

In this paper, a method has been proposed which will make it easy for the visually impaired person to surf the internet. Linking of the speech technology with the braille characters is done as these characters are the integral part of a blind person, they do reading and writing with these letters. As visually impaired person can't access the digital content more effectively as compared to normal person, speech technology is use which will dictate the content of web page to the person. Also the browser is designed in such a way that it will take the input from the user in the voice and then it will be converted to text using speech recognition algorithms, the text is send over the web server and an appropriate reply is then responded back. This process is also known as speech recognition. The reverse of it i.e. the text is then converted back to the speech using speech synthesis algorithms and then is read out to the user. The browser that performs both the above functionalities is called voice enabled browser. The major objectives of the research are as follows:

- Browsing through a Voice Recognition System.
- To propose a voice browser that allows the visually impaired learners to send and receive e-mails.
- To propose a browser that converts the Text content present over the web into Braille dots language.
- Conversion of Braille to text which will help teachers to mark assignments of the visually impaired learners.

Algorithms used for the speech recognition are HMM (hidden markov model) and DTW (dynamic time warping) and algorithms used for synthesis is CELP (code excited linear prediction) and MBROLA (Multiband Resynthesizes Overlap add Method). These algorithms are explained in the following sections [6]. Also the algorithms for converting braille characters to text and text to braille characters are discussed.

A. Speech Recognition Process

It is the process of converting any input speech signal through microphone into its respective orthographic/text representation [7]. Various phases of speech recognition are explained in fig 1.

The user speaks through the microphone or records his voice through the microphone. Microphone then sends the voice to the computer sound card after converting it the digital signal [7, 8]. Basically the analog input is converted into digital form with the help of microphone and computer sound card because computer only understand the binary language (0 and 1).

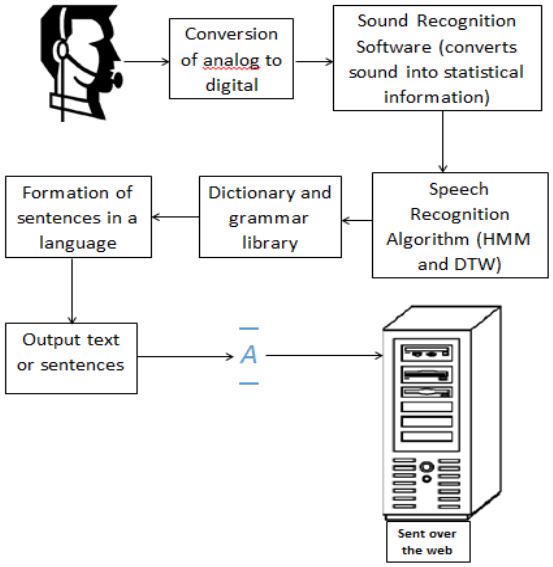

Figure 1: Conversion of speech into text

The above converted signal is then passed to sound recognition software. Sound recognition software contains acoustic models which takes input in the form of speech and create their statistical representation. Latest versions of speech technology have been designed to eliminate the noise and useless information that is not required .The words which are spoken are transformed into digital forms of the basic speech elements which are called phonemes.

Once this is complete the language is compared to the dictionary that is stored in computer memory. This is a vast collection of words, usually more than 50,000. When a match is found based on the digital form from the dictionary, that particular word is displayed on the screen [7, 8]. This is the basic process for all speech recognition systems and software. This process is done with the help of algorithms as mentioned above. The algorithms are implemented in a decoder. The detail of which is given below:-

1. Hidden Markov Model Algorithm.
   - HMM is a most widely used algorithm no- a-days.
   - It is used to find unknown parameters with the help of known parameters.
   - HMM is 92% accurate and DTW is 77%accurate.
   - HMMs are popular because they can be trained automatically and are simple and computationally feasible to use.
   - HMM is most basic and elementary algorithm for speech technology. It is more précised as compared to DTW.

Consider a system which may be described at any time as being in one of the state of set of N distinct state, U1, U2, U3, • • • , UN. At regularly time interval system undergoes a change of state (possibly back to same state) according to set of probability associated with the state [8]. Time associated with state change is denoted as t =1, 2 .., and the actual state at time t is denoted as rt. Calculate probability of occurrence by predecessor

bmn= P [rt = Um|rt−1 = Un] 1≤ m, n≤ N.

Elements of HMM

1. Number of state N
2. Number of distinct observation symbol per state M,

V = V1, V2, • • • VM

3. State transition probability,

bmn= P[rt = Um|rt−1 = Un] 1≤ m,n≤ N.

4. Observation symbol probability distribution in state j,

Bj(K) = P[Vk at t|rt = Un]

5. The initial state distribution π=πi where πi = P [r1 = Um] 1≤ m≤N [2].

Given appropriate value of N, M, A, B and π, HMM can be used as generator to give an observation sequence

O = O1 O2 O3 • • •OT

Where O1, O2… OT is observation sequences with time T.

After completion of this process, the sentences or words are identified or formed from the pre-stored dictionary or grammar library. The output is then in the form of text to the browser which interprets it in voice xml file.

B. Speech Synthesis Process

The task of speech synthesis is to convert written text that is received from the server to speech [9]. The process of speech synthesis is explained with the help from fig 2.

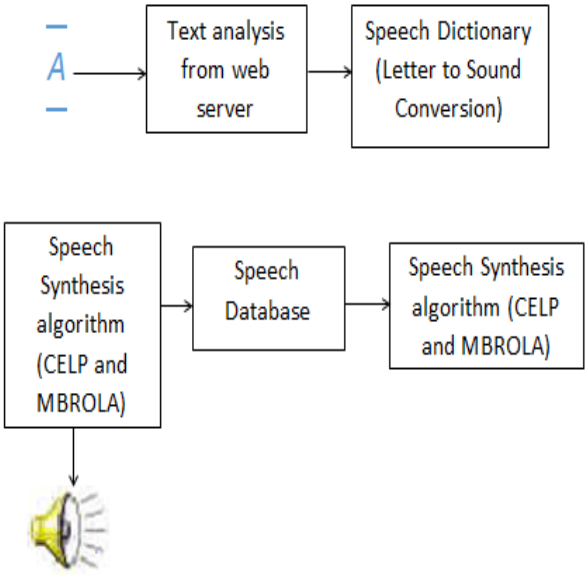

Figure 2: Conversion of text to speech

Text received from the web server is analyzed i.e. this process puts html tags into the received text [9]. This is also called normalization of text. After this speech dictionary is used for conversion of text into sound. It contains 2 processes:-

Phonetic Analysis: Text is split into phonemes and conversion to sound of these phonemes is done .Each letter is converted to sound.

Prosodic Analysis: Melody, accent, and pauses are added to the phoneme string.

Speech Synthesis algorithms are used for above said processes:-

1) Code Excited Linear Prediction (CELP) Algorithm.

Code Excited Linear Prediction is an analysis by synthesis procedure introduced by Schroeder and Atal. It is a dominant speech synthesis algorithm and used for bit rates between 4kb/s and 16kb/s. History of past excitations is stored in Adaptive codebook. The fixed codebook consists of Gaussian random numbers that are independently generated [10]. Search complexity is reduced with the help of these codebooks. The results obtained from these searches are called gains and are added with linear predictor factor to construct a speech [10]. The work of post filter is to enhance perceptual quality.

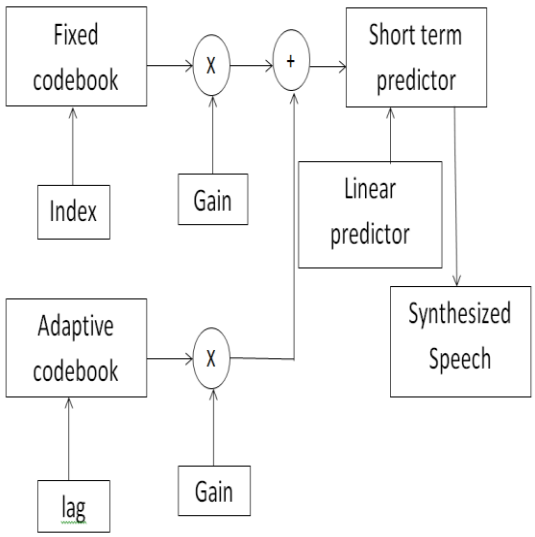

Figure 3: CELP Algorithm

After these process speech database and module is used for the exact, precise and accurate conversion into speech sound which is then heard by the listener. A web page reader program is used to dictate the content of the webpage. In this way visually impaired person is able to read the content of the web page with a great ease. Thus a speech technology plays an important role in their lives.

2) Multiband Resynthesis Overlap Add Method (MBROLA)

MBROLA is an acronym for Multiband Resynthesis Overlap Add Method. It requires minimal effort to produce a speech with high quality and accuracy. This Algorithm as compared to CELP is much improved and enhanced, which yields good performance when applied to all bit rates. Although when perceived on voice segments, there occur a certain disturbance or a slightly buzz sound which can be improved via phase control strategy. It is inspired from PSOLA algorithm. The distinguishing feature between these two is MBROLA doesn't require a preliminary marking of pitch periods. It uses diphone database according to the requirement of the synthesizers. Diphones are the elementary speech unit modules which are concatenated using MBROLA algorithm. MBROLA, indeed, is a time-domain algorithm with outstanding diphone smoothing capabilities, due to the particular format of the diphone database [11]. MBROLA is the best candidate for multi-lingual synthesis by use of diphone databases developed all over the world. Apart from this, its average computation complexity has

been kept low as 7 (operations/samples).It is a hybrid speech synthesis algorithm which is used in agriculture applications. Its process is shown in Fig:-4

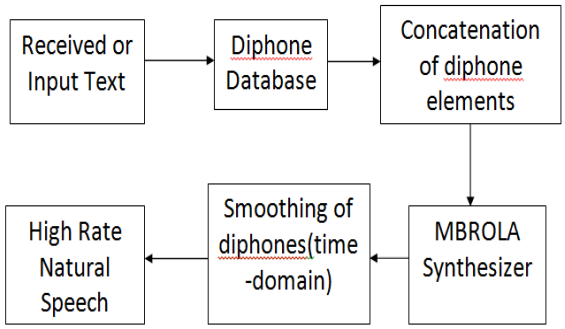

Figure 4: MBROLA Algorithm

When the visually impaired learner requires a desired article, the number is mentioned through the microphone and the system reads out the abstract of the article. If the abstract is satisfactory, and the visually impaired learner wants the full article, the learner can say print and the system will convert the text into Braille through the Text to-Braille (TTB) module for the user. Text to Braille conversion is explained with the help of fig 5:

C. Text To Braille Conversion

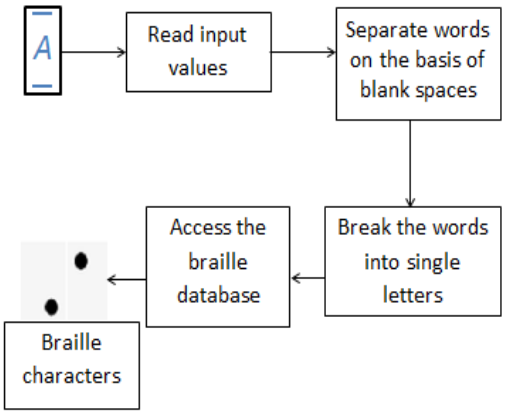

Figure 5: Conversion of text into braille characters

The steps to convert English text to Braille code are as follows:

- Input value is read till the enter key is pressed.
- Words are separated if any blank space occurs.
- Single letters are formed.
- Braille database is accessed on the basis of following conditions:

Small letter Braille Symbol is printed if the input value is between "a to z" [12]. And similarly capital letters are printed for input "A to Z". If the input value is digit i.e. 0 to 9 then it prints the Braille Numbers from the Braille Database. Suppose the input value is in special symbol list (! @#$ %^&*()_+:";'<>? []; /, etc.) [12]. Then it prints the corresponding Braille symbol from the Braille Database. Repeat the step 4 until all the characters of the input values are matched with database. If a character does not match in Braille Database then appropriate error message is generated.

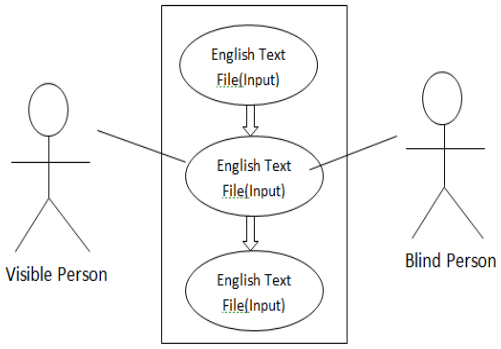

Figure 6: Braille code conversion system [6]

Figure 7: Braille databases of alphabets [13]

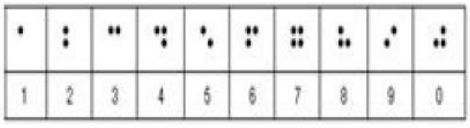

Figure 8: Braille database of numeric no's [13]

Currently, the teachers have to convert the assignments written in Braille into text manually before they can mark the assignments. This is tedious and took a lot of their time.

Thus braille to text module is used, with this module; the assignments of the students done in Braille can be automatically converted back to text on the web as well. This saves a lot of the teacher's time. Also visually impaired person can post their assignments online which can be used by various users across the world. The algorithm is explained with the help of fig 9:-

D. Braille to Text Conversion

The algorithm is just reversed of the previous one. In this The Braille character is extracted in each case and matched with the corresponding alphabet with help of a database structure.

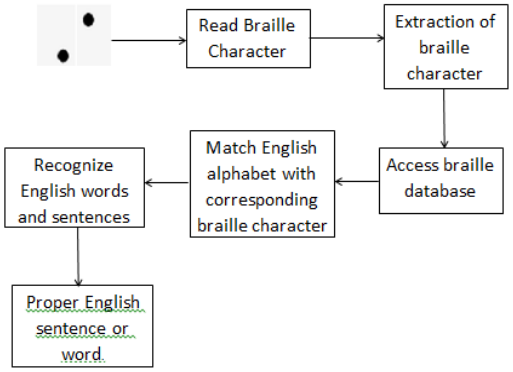

Figure 9: Conversion of Braille characters into text

For further conversion the number sequences of the Braille characters is typed with the help of number pad [14]. Matching of the number sequence is done as the numbers are typed and the corresponding alphabet is displayed when a delimiter is encountered. A voice corresponding to the alphabet is delivered as a feedback to the user. A beep is sound in case of an error. This enables the person to rectify the alphabet immediately.

Thus braille characters can be converted into simple English letters or sentences by simple algorithm. Also as mentioned before speech technology can be integrated TTB and BTT which really makes the job easy for the blind person to access the internet just like the normal person without any difficulty [14]. The above proposed methodology and modules are incorporated in a voice enabled browser or talkative browser which can also be further used to send and receive e-mails over the internet or search the content. Thus reading, writing, and surfing the net now will be easy and more comfortable from the blind perspective. These people don't have to learn any new technology or mode of communication. They just have to speak and listen to the computer and rest of the work is done by the browser. Conventional Braille characters can also be utilized in a different way as mentioned above depending upon the needs of the impaired person.

Table 1: Mapping of Alphabets [13]

Character	Representation	Character	Representation
A	7	N	7851
B	74	O	751
C	78	P	7841
D	785	Q	78451
E	75	R	7451
F	784	S	841
G	7845	T	8451
H	745	U	712
I	84	V	7412
J	845	W	8452
K	71	X	7812
L	741	Y	78512
M	781	Z	7512
:	45	;	41
"	412	!	451
,	4	.	452

4. Conclusion and Future Work

Through this paper, an efficient way of accessing the web browser is presented which is termed as voice browsing in which visually impaired people can access the browser using speech. As access to internet visually incurs limitations such as visually impaired persons cannot use keypads, touch screens etc. for giving inputs to computer. Also, teachers previously were converting braille dots into text manually, now this browser reduces their effort by performing this conversion automatically. And the blind student can also use this browser to convert text documents in braille characters. This technique allows a person to access internet services even in situations like driving etc. where user operate web just by listening and speaking rather than typing. Thus combination of browsing with speech technology is an efficient way of accessing webs. This methodology can be further improvised for a browser that allows visually impaired learners to interact more efficiently with the browser by

converting their braille characters to speech i.e. listening of braille characters, which can be easily understood by them. In addition, all the multimedia content present over the web for example images, video and animations can be made accessible by using speech technology. This technology can also be implemented in smartphones. More work can be done to increase the accuracy, pronunciation and precision of speech technology. The proposed method has used only English language but the system can be implemented in the other languages as well for e.g. Tamil, Bengali, and Thai etc. in future.

References

[1] Khushbu, Manika Kapoor, Ayesha Tafsir, "The New Era of Browsing -Voice Browsing", International Journal of Engineering and Computer Science, ISSN: 2319-7242, volume1 issue 2, page No. 59-62, Nov 2012.

[2] Suchato, Atiwong, Jirasak Chirathivat, and Proadpran Punyabukkana. "Enhancing a Voice-enabled Web Browser for the Visually Impaired." In Proceedings of the International Conference on Applied Science. 2006.

[3] K.Sireesha, A.Supriya, D.Haritha, K.S.Swetha Joseph Sastry, "Voice Recognition browser for reduced vision and vision loss Learners", International Journal of Scientific & Engineering Research, volume 2, issue 12, ISSN 2229-5518, December-2011.

[4] Edgington, M., and A. Lowry. "Residual-based speech modification algorithms for text-to-speech synthesis." In Spoken Language, 1996. ICSLP 96. Proceedings. Fourth International Conference on, vol. 3, pp. 1425-1428. IEEE, 1996.

[5] Oyama, Y., T. Tajima, and H. Koga, "Character Recognition of Mixed Convex- Concave Braille Points and Legibility of Deteriorated Braille Points", System and Computer in Japan, Vol. 28, No. 2, 1997.

[6] Dr. E. Chandra, A.Akila, "An Overview of Speech Recognition and Speech Synthesis Algorithms", International Journal of Computer Technology & Applications, Vol 3 (4), 1426-1430 ISSN: 2229-6093.

[7] S.M. Ahadi, H. Sheikhzadeh, R.L. Brennan & G.H. Freeman, "An Efficient Front-end for Automatic Apeech Recognition", published in the 2003 International Conference on Electronics, Circuits and Systems (ICECS2003).

[8] Parwinder pal Singh, Er. Bhupinder singh, "Speech Recognition as Emerging Revolutionary Technology", International Journal of Advanced Research in Computer Science and Software Engg 2 (10),pp. 410-413, October- 2012.

[9] Er.Sheilly Padda, Er. Nidhi, Ms. Rupinderdeep Kaur, "A Step towards Making an Effective Text to speech Conversion System" Vol. 2, Issue 2, pp.1242-1244, Mar-Apr 2012.

[10] Arun Kumar and Allen Gersho, Fellow, IEEE, "LD-CELP Speech Coding with Nonlinear Prediction" IEEE SIGNAL PROCESSING LETTERS, vol. 4, No. 4, April1997.

[11] T. DUTOIT, Kluwer, "An Introduction to Text-To-Speech Synthesis", Academic Publishers, Dordrecht Hardbound, ISBN 0-7923-4498-, 312 pp, 7 April 1997.

[12] Dr. B. L. Shivakumar, M.Rajasena thipathi, "English to Braille Conversion Tool using Client Server Architecture Model", International Journal of Advanced Research in Computer Science and Software Engineering 3(8), pp. 295-299, August - 2013.

[13] S.Padmavathi1, Manojna K.S, Sphoorthy Reddy and Meenakshy, "Conversion of Braille to Text in English, Hindi and Tamil Languages", International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.3, No.3, June 2013.

[14] Cheng Xu1, Ali Israr, Ivan Poupyrev, Olivier Bau, "Chris Harrison, Tactile Display for the Visually Impaired Using TeslaTouch", ACM 978-1-4503-0268-5/11/05, May 7–12, 2011.

Sachin Kapoor, currently reside in Delhi, India. I was born on March 24, 1990. I am a student and Pursuing B.Tech in Computer Science and Engineering from Bhagwan Parshuram Institute of Technology, GGSIPU University.

Adit Gupta, currently reside in Delhi, India. I was born on October 26, 1992. I am a student and Pursuing B.Tech in Computer Science and Engineering from GGSIPU University.