Voice Command and Face Motion based Activated Web Browser for Differently Abled People

LDS Tharaka[1#], DU Vidanagama[1] and WMSRB Wijayarathne[1]

[1]*Department of Information Technology, Faculty of Computing, General Sir John Kotelawala Defence University, Sri Lanka*

Abstract*: There are numerous people who are unable to use their hands due to a disability from birth or external factors like accidents. There are many military personnel who also have lost their arms and hands because of bombing and shootouts from the war in the country. These people require further assistance from another person to access the Internet. They are embarrassed because they cannot do it independently. They might hesitate to seek assistance constantly. This also raises the issue of their privacy. But because of technological advancement, now people who are disabled can do their work very easily. There are many developed applications with the voice command ability to search or type something. The main difference between those applications and the proposed system is it has the ability to use voice commands to control and use face detection to control the mouse cursor in the same application. Differently abled and handless people can use this web browser to do their work as a normal person. This proposed browser may use a voice command to control the mouse cursor and it can use voice to text to type URL. Further, computer vision may use to control the mouse cursor. This web browser can use to ease the work of normal people as well as used by disabled personnel. Both the voice command module and the web browser module were created using python. Additionally, the python Application Development Kit was used to develop the front end and back end. The researcher used open-cv to create the cursor control module. Numerous libraries have been used in the system's development. PyQT5 and QtWebEngine were both used in the development of the web browser application module. Google Speech Recognition engine API and PytTsx3 were both utilized in the development of the voice command module. OpenCV-Python-cv2, Numphy, Dilib, and autopy were used in the creation of the cursor control module.*

Keywords: *Face Detection, Speech Recognition, Assistive Technology*

1. Introduction

Voice is a very natural way to interact with a computer or device. It is still a struggle for technology in our culture, and as time goes on, we see more items and technology that can be operated by voice. Human-machine scenarios, such as those involving elderly or physically disabled people who do not have free hands to type in commands, could benefit from allowing users to command their devices or systems via voice. This would give users more control over how their devices is controlled than they could get from having someone else do it for them.

It is the goal of voice control to enable persons with special needs or those who want a luxurious and sophisticated house with a system that can understand what they say, respond to voice commands and control in real time, and achieve a hands-free computer interface[1]

When it comes to the use of face detection control over the proposed web browser it can be used to control the mouse cursor by using face coordinates gotten by the webcam. This system frees the subject from having to wear a controller to move the cursor on a computer. The subject's facial expression is recorded via the web camera, which is connected to the computer. The PC sends information about the X-Y coordinates of the cursor location while continually tracking the movement of face points. A voice browser (hardware or software) generates voice output, analyses voice input and may take and produce additional modalities of input and output, such as text input and output using vocal markup languages. Currently, voice browsers are used primarily to allow visually impaired people to speak and listen to information available on the internet via a phone, mobile phone, or any other device, hence increasing the accessibility of web-based services.

Voice browsers create synthetic speech or playback prerecorded speech as output using spoken words as input, depending on the kind of browser. There are a wide variety of hardware devices where voice browsers can be used, including telephones or cell phones; handheld computers; palm-sized computers; laptop and desktop computers; and televisions; radios; VCRS (voice-activated control systems); remote controls; ovens; coffee pots; and doorbells. In the real world, a voice browser is essential for gaining access to both public and confidential information, as well as for facilitating voice-mail communication. Currently, users interact with voice browsers only using telephones and mobile phones. Other modes and media, including pen, video, sensor, and graphical animatic and actuator controls, may potentially be supported by voice browsers in the future. If a user is functionally blind or needs online access, they can utilize voice browsers and voice controllers to display an unseen user interface to the

user while keeping their hands and eyes free for other tasks, thereby freeing up space previously taken by keyboards and mouse.

"Devices which interpret (voice) markup languages and are able to generate voice output and/or interpret speech input" are known as "voice browsers." A general definition of a voice browser is provided above. Given that "browser" appears first in its name, it is clear that the system handles speech. However, what makes it a "browser"? When it comes to either domain data or dialog flow, the system relies on information from the internet. The goal is to provide a service comparable to what graphical browsers of HTML and associated technologies do today, but on devices that do not have full browsers or even displays large enough to handle them. The problem is made worse by the fact that a lot of today's content relies on the correct functioning of scripting languages and third-party plug-ins.

Machines and programs with voice control can take dictation or recognize and execute spoken commands. A computer system may also be controlled by a voice or a series of voices that the computer has been trained to recognize. The structure of this paper is as follows, initially reviewed the literature on assistive technologies that have been used to prevent the uneasiness of differently abled personals. Then conducted the survey to study the background of differently abled handless people and how they use the current system to browse the internet. Then identified the research gap. The Voice Command and Face Motion-based Activated Web Browser have been proposed as the solution.

The proposed web browser can be used by differently abled handless people who are unable to operate the current web browsers. The proposed web browser will be able to operate through voice commands and move the mouse cursor using face motion. The Objectives of the proposed web browser are identifying the problems faced by differently abled handless people when browsing internet. When developing the proposed web browser identify the relevant technology to establish the functional and non-functional requirements are important. Most crucial factor is in this research design and develop the voice and face detection-based web browser that could be used by differently abled persons. Finally the proposed system should test and it should evaluate. The present work aims to introduce a web browser that could be operate by differently able handless people using their voice and face detection to give command and control the cursor. The goal of this research is to learn about the limitations, obstacles, and maturity level of voice control and voice browsers, and how speech recognition may be enhanced to make opportunities more accessible to those with physical disabilities and to those who desire a luxurious lifestyle. How voice prompts can be used to control digital devices is also discussed, as well as suggestions for how to make voice accessibility even better. However, it is more about how those who are physically disadvantaged may adapt to new technology that are already or will be a part of their daily lives[2]

Furthermore, the possible solutions to these limitations, such as restricting users to specific commands and how to deal with the difficulties posed by cultural (and time zone) and linguistic diversification, are also discussed. People from various ethnic groups and nations may have difficulty with the same voice command, which is often an English term, because of the variances in their pronunciation. Also explored are prospective future uses for voice control and voice browsers, as well as how we can use this technology on a day-to-day basis soon.

2. Literature Review

A.History of Voice control and Voice browsers.

In 1936, AT&T's Bells laboratories began studying automated voice recognition and transcription. It was not until the early 1980s that the technology became available to the public, financed mostly by the military and DARPA the Defense Advanced Research Project Agency.

When Covox launched its first commercial product in 1982, it was the first firm to do so. In the 1980s, Covox introduced digital sound to the Commodore 64, Atari 400/8, and IBM PC with the voice master, sound master, and speech thing. Speech recognition was introduced together with (or as part of) the introduction of sound to the computer. It is also worth noting that Dragon Systems was created in 1982 and has gone on to dominate the voice recognition market with their software. ScanSoft, Inc.

For many years, Audrey was a speech-recognition system built by Bell Labs. It was a simple system that only recognized the first nine digits of the alphabet. It also necessitated a pause between each syllable, making it difficult to utilize. IBM's "shoebox" gadget, introduced in the early 1960s, could understand 16 complete phrases, including 10 numbers and 6 arithmetical commands. Both Audrey and Shoebox were, of course, quite impractical by today's standards because they could not be easily transported[3]

In the 1970s, the Department of Defense became interested in voice recognition and contributed some funds for research, which sparked the development of Harpy. The "Harpy" system, developed at Carnegie Melon shortly, thereafter, had a vocabulary roughly equivalent to that of a child, with a substantial increase over earlier systems. Speakers still had to stop between words so that the computer could identify them, which meant that this technology was severely constrained in terms of more than just vocabulary. Speech recognition truly took off after Harpy when the hidden Markov model (HMM) was introduced, which would later be used by IBM, Philips, and Dragon systems to build their voice recognition systems.

To expand the vocabulary of a voice recognition system, HMM enables systems to take into consideration the

potential that unfamiliar sounds may constitute words. For example, the Julie Doll by worlds of wonders "The doll that knows you" made it feasible for voice recognition technology to find commercial uses. Dragon was born in the 1990s when computers grew in popularity and processing power exploded, making computers a viable commodity.

It was the first consumer voice recognition system to be sold for more than $900, and it was called Dragon Dictation. With Dragon Naturally Speaking (DNS) in 1997, you no longer must stop and think about what you are saying before continuing. Still $645 and requiring a lot of time for the consumer to learn, despite a price reduction.

After the creation of Dragon, Siri saw that things had slowed down a bit. Though a few people were aware of the feature, both Vista and Mac OS X featured voice recognition incorporated into the operating systems.[2]

To put it another way, voice recognition was going to be one of those novelty technologies that never found its place in the mainstream. Then, in 2007, "Speakeasy" was introduced. As a follow-up, Google voice search was launched for the iPhone in 2008. In addition to the mobile interface being great for voice recognition, as we had discovered the year before with "Speakeasy," the app had access to more computational power than ever before by shifting the data processing essential for speech recognition to Google's cloud data centres.

New personalization features were implemented in 2010, allowing the program to learn from you and become more accurate than ever before in 2011, "Siri" was debuted, and it rapidly became a sensation. For better transcription accuracy, it makes advantage of cloud-based processing, as well as artificial intelligence and personality. There's little doubt that speech recognition has reached the public, despite a fall in usage after the novelty wore off.

B. History of face detection

Woody Bledsoe, Helen Chan Wolf, and Charles Bisson were among the first pioneers of face recognition. Bledsoe, Wolf, and Bisson started working on computer-based facial recognition in 1964 and 1965.Much of their work was never released since the research was sponsored by an unnamed intelligence agency. But it was later discovered that their early effort involved physically marking different facial "landmarks" like the mouth and the centres of the eyes. To compensate for pose variation, these were then mathematically rotated by a computer. In order to identify landmarks, the distances between them were also automatically calculated and compared across images[4]

These early experiments in facial recognition by Bledsoe, Wolf, and Bisson were greatly limited by the technology of the time, but they still represent a significant first step in demonstrating the feasibility of facial recognition as a biometric.

Following up on Bledsoe's original research, Goldstein, Harmon, and Lesk continued the study in the 1970s and expanded it to include 21 distinct subjective markers, including hair colour and lip thickness, in order to automate recognition. Even though the precision was improved, the measurements and positions had to be calculated manually, which proved to be incredibly labour-intensive but still a development over Bledsoe's RAND Tablet technology.

The Facial Recognition Grand Challenge (FRGC), which was established in 2006, has as its main objective the development and advancement of face recognition technology intended to assist ongoing face recognition initiatives within the U.S. Government.

The most recent facial recognition algorithms were assessed by the FRGC. The experiments made use of high-resolution facial pictures, 3D face scans, and iris photos. The results demonstrated the improvements in facial recognition technology over the previous ten years, with the new algorithms being 10 times more accurate than face recognition algorithms from 2002 and 100 times more accurate than those from 1995.

Social media started adopting facial recognition technology in 2010 to assist in identifying persons whose faces could appear in the pictures that users upload on a regular basis. The news media immediately found the feature to be contentious, which led to a flurry of pieces on privacy

C. Uses of Voice control and Voice browsers

Voice control is an option for customers that require quick access to basic information. A live operator may not be necessary or desired in many situations. Speech recognition, for example, can be used to shorten wait times and offer consumers with the information they need if they only have a limited amount of time or information needs. An intelligent speech system allowed Dublin Airport to handle a 30% increase in passenger traffic without increasing the number of staff members. When a consumer phones in, the system immediately routes them to a voice recognition system, which uses a series of questions to determine whether they need information on "departure" or "arrival."

Phone calls may be routed to the appropriate department using voice commands. Customers may become irritated and dissatisfied if they must wait in a long line to speak to an operator. Using voice recognition, may lead customers to "self-service" options or just "speak" what they need, and the appropriate department or person will be contacted. To confirm someone's identity on the phone, voice commands can be utilized instead of "risky" personal data. Identity fraud is a huge worry for some organizations, and a sophisticated speech recognition system gives a response to this problem utilizing voice biometrics, which is a crucial instrument in preventing telephone-based crime4[5]

When your phone's call volume surges or you need to navigate your Android phone or tablet, you may utilize voice controls to manage the situation. Some voice browsers may be taken on the go. There is no limit to where you may utilize them. More individuals will be able to access information, especially those without networked computers but with access to handsets, such as cell phones

D. Applications of Voice browser and voice control

Other applications include teaching students of other languages how to pronounce words correctly and translating across languages, as well as controlling autonomous robots in a multi-robot system via voice control. Appliances like the oven, refrigerator, and dishwashing machine may all be controlled using voice commands.

Vocal control can be used in military aircraft and helicopters as well as in combat management systems for teaching air traffic controllers. Medical transcriptions and health care may both benefit from the use of voice control in the industry (digital speech to text).

E. Early Improvements

Text to speech and speech to text technologies were successfully merged and used as a communication channel between two physically challenged people in this Speech Recognition & Synthesis Tool. System. To turn spoken words into text, a Speech. Recognition engine is employed. The W3C standard for voice recognition is supported by SAPI 5.3. The SRGS markup language, which specifies how and what words are recognized by the engine, is also supported[6][7]

Using speech recognition technology to complete the university's written proficiency exam proved to be a beneficial trial at California State University in Northridge. They were able to obtain an equal distribution of exam marks with their nondisabled counterparts thanks to this innovation[8]

A sixth-grade kid with learning impairments was the only focus of another exploratory investigation. For Wetzel, the goal was to see if middle school children could learn to utilize a speech recognition system, such as IBM Voice Type, and whether this technology would help them improve their communication abilities. Although Wetzel observed that the student was able to grasp the program, challenges with the system's recognition accuracy and the intricacy of editing hampered this student's performance[9]

The Frostig Center for Learning Disabilities in Pasadena, California, researcher Marshall Raskind revealed that speech recognition software might make a significant impact for many dyslexics. Speech recognition can help dyslexics speak more effectively, but it may also help them overcome their disease, according to a new study[10]

A voice-activated browser and screen reader are included in a specialized small web browser. According to the GNU General Public License, the website browser is a free software project. Using the text-to-speech engine, the new built-in screen reader may now be enabled by just hovering the mouse pointer over it. In addition to the spoken command input, the browser now has a dialog module. Web page navigation is possible with this tool. To utilize the new online site, which is reserved for blind and visually impaired people alone, the Developed Browser has been designed specifically. It is expected that all sites on this portal and those connected to it will conform to the Web Access Initiative's fundamental standards for HTML/XML pages[11]

A cloud-based wheelchair platform that controls prototype speech is described in this article. Cloud-based speech WebKit is under the control of the platform. It also works on online browsers and mobile devices that broadcast live video, in addition to voice control. The Arduino UNO Microcontroller and the Mini PC running Ubuntu Linux constitute the foundation of the platform. JavaScript and ECMAScript were used to construct the software[12]

Drowsiness may be detected in real time according to the system's architecture. The program is written in C++ and runs on Windows utilizing the OpenCV library and a single camera view. Under ideal lighting circumstances, the video surveillance system technologies they described offer promising outcomes[12]

In many circumstances, people with neuro-Locomotor difficulties are able to comprehend and should communicate through their eyes. Eye-tracking mouse-based system is provided in this study, which is both scalable and cost-effective. When the user's eyes move, a headmounted device picks up the movement and moves the mouse pointer on the screen. If the patient gazes at the relevant picture on the screen for a predetermined amount of time, a click event signifying a pictogram selection is executed[11]

3. Proposed Methodology

There are many military personnel who also have lost their arms and hands because of bombing and shoot outs from the war in the country. These personnel need the help of another person to access the Internet. They are embarrassed because they cannot do it independently.

They may feel uneasy to always get help. This also raises the issue of their privacy. But because of the technological advancement, now people who are disable they can do their works very easily.

Due to various reasons, there are many people who are with disabilities, and they should totally rely on others to access the Internet. So now there are already developed applications with voice command ability to search or type something. But there is no customized web browser

developed for differently abled and handless people. In order to address this issue, The researcher proposed a web browser that can use without hands. differently abled and handless people can use this web browser to do their works as a normal person.

As a methodology for conducting this research first studied literature that considers assistive technologies that has used to prevent uneasiness of differently abled personals. Then conducted the survey to study the background of differently abled handless people. Then researcher has identified the research gap.

According to the literature survey above conducted it is evident that control a web browser for differently abled and handless people is much needed implementation since only few research has been done targeting some few parts of it. Up to now in the research that was conducted under domain specific that have covered following domain

.

i. physically disability services
ii. improve assistive technology for learning disability
iii. Voice driven and speech recognize
iv. Detecting eye movement by computer vision

Specifically, the problem identified in this scope is use the internet by differently abled handless people. Above research have covered some parts in this problem domain but there is no specific solution for this problem. Below table shows the comparison of above research and proposed system.

Table 1. Comparison of the existing system and the proposed system

Research	**Web browsing application for differently abled people**	**Computer vision-based solution**	**Voice drive n solution**	**Integration of modules to specify the problem identified**
(Sharma and Wasson, 2012) [8]	No	No	Yes	No
(Higgins and Zvi, 1995) [10]	No	No	Yes	No
[11] [11]	Yes	No	Yes	No
(Dhaval Pimplaskar Atul Borkar, 2013) [13]	No	Yes	No	No
Proposed system	Yes	Yes	Yes	Yes

According to the conducted survey 14 personals have answered to the questioner that who cannot use both hands and disabilities in their hands. According to the survey 78.6% of differently abled handless people are like to use internet like other normal people.71.4% has feel uneasy when they are using internet by assistant of a third-party person.78.6% people is needing a new solution that they can use internet without the help of a third party.

A. Approach

In the Analysis of the Voice command and face detection-based web browser people can use the web browser for browse the internet by only using voice commands and face detection to control the cursor. User can type URL on the URL bar in the web browser by speaking though the microphone and turn spoken words into text, and the user can type the specific URL of the web site that user needs to browse. When it comes to the search functions in the search engines user also can use the voice to text feature in the proposed browser. Use can use the other functions in the web browser by giving voice commands. When user need to open a tab, close the current tab, go backward through web pages, go forward through web pages, refresh the current web page can use necessary voice commands featured in the proposed web browser. Apart from the voice command one of the main functions in this web browser is the user can control the mouse cursor by using face detection and user can move the mouse cursor by moving the face. In the mouse scrolling, clicking function can use by using the voice commands.

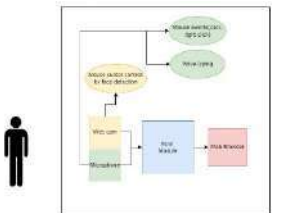
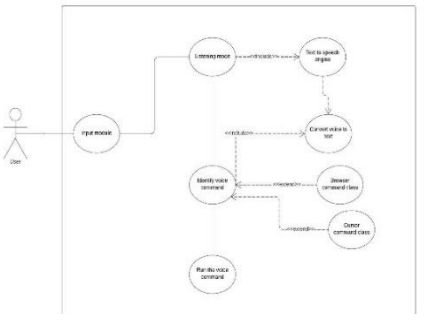


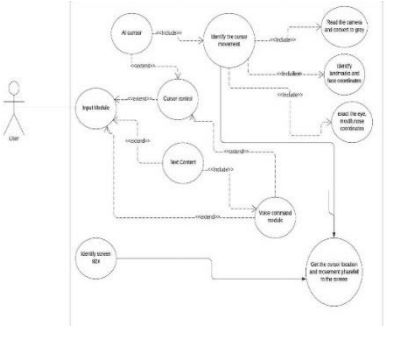
Figure 1 Block diagram of the system

.



.

Figure 2 Use Cases Diagrams Cursor control module



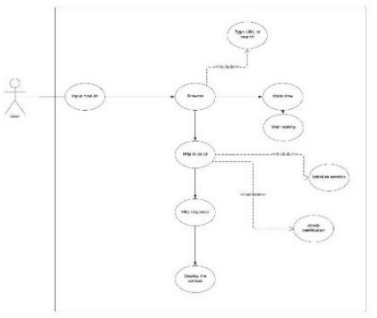


Figure 4 Use Cases Diagrams web browser module

B. Technology Adopted

The system should use the newer technological methodologies. It is important to use necessary tools.in order to create a productive system. Use of any inappropriate Technologies may only result in the development of a system having errors and faults. Badly chosen technologies additionally can end up in crashed when the new system implementation. Badly chosen technologies which can be extremely advanced and complicated will enable manufacturing a system with a top quality, however, these technologies may result in developing a system that spends lots of time and resources to perform a task that is anticipated by the system. The most important goal of creating this kind of application is to give consumers a more efficient work system. Technologies and tools can assist to develop the system with a minimum amount of time. Because of that, should use the most applicable tools available to develop the system.

Technological considerations followed during the development of the system Efficiency and Performance Re-usability and flexibility object-oriented development support so according to the proposed voice command and face detection-based web browser python has used to develop the web browser module as well as the voice command module. The programming language that will be used to build the system, acquired the trust, accuracy, and efficiency. When considering all these technologies which can be associated with the proposed system can be applied python application building tool. As well as the front-end back end also developed with the python application development kit. To develop the cursor control module researcher has used open-cv. There are many libraries have used to build the system. To develop the web browser application module PyQT5 has used and as the web engine QtWebEngine has used. To develop the voice command module PyWhatKit, PytTsx3 and google Speech recognition engine API has used. To develop the cursor control module OpenCv-Python-cv2, Numphy ,Dilib and autopy has us

When discuss about the functional requirement proposed web browser can control the cursor with the face movement. In order to type an URL or search on a search engine, proposed web browser can use voice to text to enter URL and Search in internet. Proposed web browser can use voice command to control the web browser events such as clicking and scrolling events in browser functions.

When discuss about the nonfunctional requirement proposed web browser the Efficiency is one of important attribute the proposed web browser performance take minimum 512RAM and disk Space - minimum 1GB. When discuss about the Security, differently abled handless people should be able to browse internet without the support of third party and the interoperability of the web browser should be able to use in different kind of operating systems. When talk about the usability of the proposed system the web browser should be able to use easily perform task, effectively, and efficiently while enjoying the experience.

4. Discussion

The results and outcomes engendered in relative to the specificity of the problem domain are enlarged into wider concepts depending on logical assumptions. the outcomes and findings of the project and to determine the way of these outcomes and findings can be matched in different contexts that are like the problems which are solved by the developed voice command and face detection-based web browser. The research problem identified in this research is People who are differently abled and handless that cannot control the mouse and keyboard must face many difficulties when browsing the internet. They face many difficulties such as inability to type a URL, using the search, browsing social media, etc. It is necessary for another person's help when doing these tasks. This may result in privacy issues since confidential passwords and user data may be needed to be shared with third parties. There is also the issue of inability to access the internet at their own leisure since they have to wait on a third party being present in the location. According to the literature review and the survey research gap has identified. The studies mentioned above literature review covered some parts in this problem domain but there is no specific solution for this problem. To fill the research gap proposed solution has developed accordingly. In this proposed browser using voice command to control the mouse cursor events and it can use voice to text to type URL. In further computer vision use to control the mouse cursor and users can move the cursor by using face movements. This is an interim presentation of an

ongoing research and no experiments or implementations conducted.

REFERENCES

[1] B. Šumak, M. Špindler, M. Debeljak, M. Heričko, and M. Pušnik, “An empirical evaluation of a hands-free computer interaction for users with motor disabilities,” *J. Biomed. Inform.*, vol. 96, no. July, p. 103249, 2019, doi: 10.1016/j.jbi.2019.103249.

[2] E. C. Matel, H. Aquines, C. B. S. de Guzman, J. I. V. Gustilo, G. V. P. Ibera, and P. D. Jodilla, “ECHO: Hands-free computer interaction using speech recognition system for the debilitated,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 2, pp. 1768–1772, 2019, doi: 10.35940/ijrte.B1017.078219.

[3] J. Babu and C. Sekharaiah, “A Panorama of Web Accessibility,” *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 8, pp. 311–317, 2014.

[4] “The History of Information Security: A Comprehensive Handbook,” *Libr. Hi Tech*, vol. 26, no. 4, pp. 682–683, 2008, doi: 10.1108/07378830810920987.

[5] J. A. Jacko, “Human-Computer Interaction: HCI Intelligent Multimodal Interaction Environments - 12th International Conference, HCI International 2007, Proceedings,” vol. 4552 LNCS, no. PART 3. 2007.

[6] S. Bhattarai, “Hands-free Text Editing using Voice and Gaze Master ’ s Thesis,” no. December, 2019.

[7] F. Authors, “Efficacy of Assistive Technology Interventions,” p. i, 2015, doi: 10.1108/s2056-769320150000001013.

[8] F. R. Sharma and S. G. Wasson, “Speech recognition and synthesis tool: assistive technology for physically disabled persons,” *Int. J. Comput. Sci. Telecommun.*, vol. 3, no. 4, 2012.

[9] K. Floyd, “Postsecondary students with learning disabilities: Can we do more?,” *J. Spec. Educ. Apprenticesh.*, vol. 1, no. 1, pp. 1–13, 2012.

[10] E. L. Higgins and J. C. Zvi, “Assistive technology for postsecondary students with learning disabilities: From research to practice,” *Ann. Dyslexia*, vol. 45, no. 1, pp. 123–142, 1995, doi: 10.1007/BF02648215.

[11] S. Dobrišek, J. Gros, B. Vesnicer, F. Mihelič, and N. Pavešić, “A Voice-Driven Web Browser for Blind People,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2448, pp. 453–459, 2002, doi: 10.1007/3-540-46154-X_65.

[12] T. Brøndsted and E. Aaskoven, “Voice-controlled internet browsing for motor-handicapped users. Design and implementation issues,” *9th Eur. Conf. Speech Commun. Technol.*, pp. 185–188, 2005, doi: 10.21437/INTERSPEECH.2005-99.

[13] M. S. N. Dhaval Pimplaskar Atul Borkar, “Real Time Eye Blinking Detection and Tracking Using Opencv,” *J. Eng. Res. Appl.*, vol. 3, no. 5, pp. 1780–1787, 2013.