A SURVEY ON AI POWERED PERSONAL ASSISTANT

Mrs. Ashwini Bhamre[1], Arpita Moharir[2], Vaishnavi Kulkarni[3], Rashmi Modgi[4], Ninad Kulkarni[5]

[1]*Assistant Professor, Dept. of Information Technology, P.E.S. Modern College of Engineering, Pune, India*
[2] *U.G. Student, Dept. of Information Technology, P.E.S. Modern College of Engineering, Pune, India*
[3] *U.G. Student, Dept. of Information Technology, P.E.S. Modern College of Engineering, Pune, India*
[4] *U.G. Student, Dept. of Information Technology, P.E.S. Modern College of Engineering, Pune, India*
[5] *U.G. Student, Dept. of Information Technology, P.E.S. Modern College of Engineering, Pune, India*

---***---

Abstract - Web accessibility is vital for inclusivity among visually impaired individuals. This survey paper delves into designing a web application tailored for their enhanced usability. Through voice-driven interfaces and assistive technologies, the objective is seamless interaction and navigation. Exploring challenges faced by visually impaired users accessing web applications, the paper emphasizes auditory feedback and adaptive interfaces. It details a user-friendly login/signup process with voice recognition, providing auditory cues. The home page offers crucial functions like battery status, search engines, weather updates, and YouTube accessible via voice or buttons. Addressing privacy, error handling, and user feedback, the paper advocates inclusive design and assistive tech integration. Drawing from academic studies, it consolidates best practices for developers, fostering an inclusive digital environment for visually impaired users.

Key Words: **Web accessibility, Visually impaired, Voice-driven interfaces, Voice recognition.**

1.INTRODUCTION

The recent evolution of virtual assistants (VAs) marks a significant shift in user interaction and experience within the realm of technology. These VAs, primarily voice-based, have transformed how individuals engage with devices, enabling tasks like controlling lights, playing music, and conducting voice-based searches. They operate through stages: text-to-speech, text interpretation, and intention-to-action, constantly evolving with AI advancements.

Studies reveal the growing reliance on VAs; a survey in 2015 highlighted that a substantial percentage of both teens and adults in the U.S. anticipated the widespread use of voice search in the future. Foreseeably, enhanced speech recognition, empowered by robust internet connectivity and advanced data processing capabilities, will augment VA capabilities, making them more intuitive and efficient.

However, the integration of VAs into daily life raises questions about user preferences and concerns. While many anticipate increased VA usage, there's hesitancy among some individuals or limitations in their usage. Limited research has explored the factors influencing VA acceptance comprehensively, especially regarding various configurations and user requirements.

To delve deeper into user preferences, a study proposes employing conjoint analysis—a method to gauge user preferences—regarding three specific aspects: natural language processing (NLP) performance, pricing, and privacy concerns. By weighing these factors against each other, the aim is to unveil user segments and their trade-offs in accepting VAs.

Simultaneously, the integration of smart assistants into mobile technology is gaining prominence due to its ease of access and versatility. A smart assistant is essentially a virtual assistant implemented on personal computers, facilitating voice or keyboard inputs and internet-based remote access. Comprising components like speech-to-text converters, query processors, and text-to-speech converters, these systems leverage voice's efficiency in communication and are integral to computer devices.

Moreover, the digital landscape's transformation has seen voice searches surpassing text searches, with predictions indicating that half of all searches will be voice-based by 2024. Virtual assistant software, designed for desktop usage, aids in managing tasks, providing information, and improving user productivity. Recent research focuses on recognizing human activities through voice recognition systems powered by Natural Language Processing (NLP) algorithms, aiming to understand user commands and respond accordingly.

Overall, the evolution of VAs, their potential impact on user interaction, and the exploration of user preferences regarding acceptance and usage constitute significant areas of current research and development in technology.

2. LITERATURE REVIEW/SURVEY

The referenced papers collectively illustrate the progressive advancements in speech recognition technology and the wide-ranging applications of voice-controlled assistants. These innovations encompass diverse areas, from personal voice assistants to smart home automation and aiding visually impaired individuals.

One notable project, 'JARVIS', amalgamates Artificial Intelligence with Google platforms, employing markup language to convert text formats into speech. Emad S. Othman's project focuses on a personal voice assistant utilizing Raspberry Pi, demonstrating its architecture configuration and efficiency in managing various tasks.

Ankit Pandey's smart voice assistant showcases functionalities such as note-taking, email exchanges, and calendar scheduling via speech commands, also extending its utility for appliance monitoring. Subash. S implements an AI-based virtual assistant adaptable for desktops and mobiles, translating spoken content into human-readable data. Yash Mittal et al.'s Smart Home Automation system utilizes Arduino microcontrollers to process voice commands for domestic appliances.

Moreover, Rahul Kumar's power-efficient smart home with a voice assistant focuses on aiding visually impaired individuals through Raspberry Pi-based hardware. Jianliang Meng et al. provide an overview of Speech Recognition Technology, emphasizing the machine's ability to comprehend vocal input through recognition patterns and signal processing.

Deepika Sherawat implements a voice-activated desktop assistant, featuring in-built commands for user queries. Prof. Emad S. Othman's Voice Controlled Personal Assistant serves as a surveillance model, catering to blind individuals and predicting information from spoken input.

Additionally, N. Vignesh's Comparative Study on Voice Based Chat Bots highlights challenges in chatbot versatility and proposes an ontology-based approach. Ankit Pandey, Vaibhav Vashist, Prateek Tiwari, and Sunil's project focuses on email sending, to-do lists, and web service tasks, aiming to integrate their voice assistant with cloud infrastructure for multi-user functionality.

3. METHODOLOGY

Voice assistants operate through programming languages, enabling them to interpret verbal commands and provide responses based on user queries. In this context, Python programming language serves as the foundation for building AI-based Voice Assistants. For instance, a command like "Play a song" or "Open Facebook website" triggers the assistant to respond accordingly by executing the requested action.

The functionality of the voice assistant involves several steps:

1. Listening for pauses after a request to understand when the user has finished speaking.

2. Parsing the user's request by breaking it down into distinct commands for better comprehension.

3. Comparing these commands with available requests stored in the assistant's database.

4. Executing the appropriate action based on the recognized command.

5. Seeking clarification from the user if the request is ambiguous, ensuring accurate interpretation.

Regarding the Automatic Speech Recognition (ASR) system, it functions by recording speech, creating a wavefile, filtering background noise, normalizing volume, segmenting the speech into elements, and applying statistical probability analysis to convert it into text content. The ASR process involves Acoustic Modeling, Pronunciation Modeling, and Language Modeling, which collectively aid in recognizing and decoding spoken elements into comprehensible text commands.

This AI-based system processes recorded speech data independently, using Artificial Intelligence (AI) algorithms without human intervention. The processed speech waveforms are transmitted to the decoder, ultimately transforming them into text commands for further execution.

Voice assistants operate on a sequence of sophisticated algorithms that enable them to comprehend and respond to user voice commands. The backbone of these systems lies in programming languages such as Python, providing a framework for developing AI-based voice assistants. These assistants are capable of executing diverse tasks, ranging from playing music or opening websites to managing daily tasks like setting reminders, sending emails, or fetching information.

The operation of a voice assistant involves a multi-step process:

1. Voice Input Interpretation: Upon receiving a verbal command, the assistant utilizes various techniques to parse and interpret the spoken words. It leverages natural language processing (NLP) algorithms to understand the intent behind the user's query accurately.

2. Command Execution: After parsing the input, the system matches it with predefined commands or tasks stored in its database. If a match is found, the assistant executes the corresponding action or command.

3. Contextual Understanding: Sophisticated voice assistants are designed to understand context. They consider previous commands, ongoing conversations, or user preferences to deliver more personalized and accurate responses.

4. Learning and Adaptation: AI-powered voice assistants employ machine learning algorithms to continuously learn

from user interactions, improving their accuracy and efficiency over time.

Regarding Automatic Speech Recognition (ASR) systems, their functionality is crucial in the voice assistant's ability to interpret and process spoken commands. ASR involves several stages:

- Audio Capture: The process begins by capturing audio through a microphone or a similar source.

- Preprocessing: Recorded audio undergoes initial processing steps, including noise reduction and volume normalization, to enhance the clarity of the voice signal.

- Feature Extraction: Speech signals are converted into a mathematical representation, typically using techniques like Mel-frequency cepstral coefficients (MFCCs), to extract relevant features for analysis.

- Acoustic Modeling: This phase involves modeling sound units and their variations to determine possible words or phrases corresponding to the extracted features.

- Language Modeling: Analyzing word sequences and predicting the likelihood of specific words or phrases occurring together, based on statistical language models.

- Decoding: Finally, the system decodes the processed speech data into text, enabling the voice assistant to understand and act upon the user's command accurately.

These systems continuously evolve with advancements in AI, machine learning, and natural language processing, contributing to their increased accuracy and capabilities in understanding and responding to human speech.

4. EXISTING TECHNOLOGIES

4.1 Accessibility Standards and Guidelines (WCAG):

A) Principles of WCAG: WCAG is structured around four core principles: Perceivable, Operable, Understandable, and Robust (POUR). These principles are further detailed into specific guidelines and success criteria to ensure the accessibility of web content for all users, including individuals with disabilities.

B) Challenges in Implementation: While WCAG offers comprehensive guidelines, there are challenges in fully implementing them. Factors such as the complexity of interactions in modern web applications, dynamic content, and the rapid evolution of technologies pose difficulties in adherence to existing guidelines.

4.2 Natural Language Processing (NLP):

A) Components of NLP: NLP encompasses several elements, including tokenization, syntactic analysis, semantic analysis, named entity recognition, and machine learning models such as transformers (e.g., BERT, GPT). These components collaborate to comprehend and process human language.

B) Resolving Ambiguity: One of the primary hurdles in NLP involves addressing ambiguity in natural language queries. Techniques such as context-aware language models and entity disambiguation are utilized to enhance accuracy in interpreting user intents.

4.3 Machine Learning and Voice Recognition:

A) Algorithms for Voice Recognition: Advancements in machine learning, notably deep learning methods like Convolutional Neural Networks (CNNs), RNNs, and transformer models (e.g., BERT), have significantly bolstered voice recognition accuracy. Strategies like transfer learning are employed to optimize models, particularly when dealing with limited data.

B) Addressing Biases: Overcoming biases in voice recognition models is pivotal for creating fair and inclusive systems. Approaches such as bias detection, data augmentation, and curated diverse datasets are explored to mitigate biases.

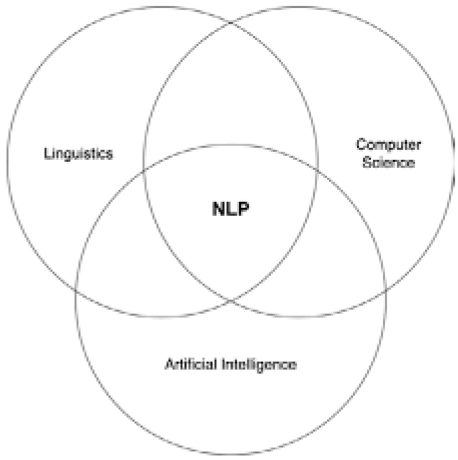

Fig -1: NLP Model

4.4 Multi-Modal Interfaces:

A) Challenges in Integration: Designing interfaces that seamlessly amalgamate multiple interaction modes (voice, touch, gesture) necessitates thoughtful consideration of user preferences, environmental contexts, and technological capabilities across various devices.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 01 | Jan 2024 www.irjet.net p-ISSN: 2395-0072

B) Consistency and Usability: Ensuring uniform and user-friendly interactions across different interaction modes presents challenges. Incorporating user testing and iterative design processes are essential for refining multi-modal interfaces, ensuring enhanced usability and accessibility.

5. PROPOSED TECHNOLOGIES

5.1 Voice User Interface (VUI) and Speech Recognition:

VUIs enable users to interact naturally with technology using spoken language, significantly improving accessibility for visually impaired individuals. This hands-free method offers an intuitive approach to accessing information and performing tasks. Unlike traditional interfaces requiring physical input, VUIs allow users to control devices or applications without manual interaction, especially beneficial for those with limited motor abilities. Accurate speech recognition technology facilitates faster navigation of applications, web browsing, and task execution compared to conventional input methods, potentially boosting productivity for visually impaired users.

5.2 Assistive Technologies:

Screen readers and text-to-speech systems deliver instant auditory feedback, granting visually impaired users' immediate access to digital content as it appears on screens, eliminating accessibility barriers. These technologies empower visually impaired individuals to independently access digital content, fostering autonomy and reducing reliance on external support. Assistive technologies often offer adjustable settings, such as speech rate, voice options, and navigation preferences, accommodating individual user needs and preferences.

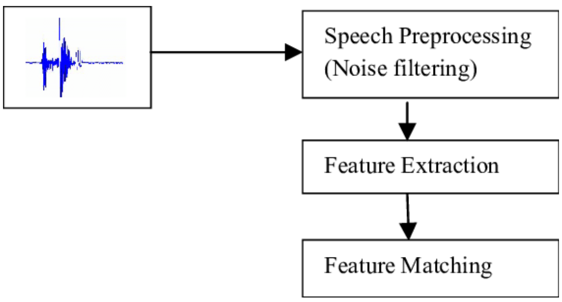

Fig -2: Speech Recognition Model

5.3 User-Centric Design and Human-Computer Interaction (HCI):

HCI principles in user-centered design ensure interfaces cater specifically to various visual impairments, considering factors like contrast, font size, and compatibility with screen readers, enhancing accessibility. User-centric design emphasizes usability testing and feedback integration, resulting in interfaces that are intuitive, easy to learn, and navigate for visually impaired users. HCI studies advocate for inclusive design strategies, considering a diverse range of users' abilities and needs, aiming to create universally accessible interfaces accommodating a wide spectrum of users.

In summary, these proposed technologies offer improved accessibility, natural interaction methods, user empowerment through independence, and interfaces tailored to meet the unique requirements of visually impaired individuals. Their integration signifies continued efforts to bridge accessibility gaps in digital interactions, promoting inclusivity and equitable access to information and services.

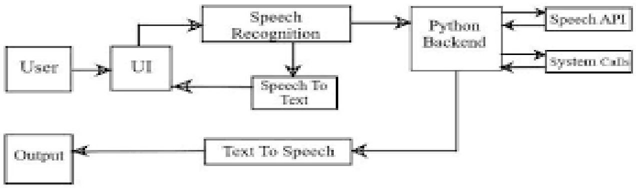

Fig -3: Proposed Model

6. RESULTS

The final outcome of the project intends to introduce a fully operational web application tailored specifically for individuals with visual impairments. This application will feature an intuitive login interface that incorporates voice-based inputs for entering user credentials such as name and phone number, thereby facilitating a smooth login or signup process. Upon successful authentication, users will gain access to a user-friendly homepage offering multiple functionalities. These include checking battery percentages and a versatile "Ask Me" feature, enabling voice-activated queries for Google/Wikipedia searches, weather updates, news, jokes, and access to YouTube content. By integrating Voice User Interface (VUI) technologies, assistive tools like screen readers and text-to-speech systems, and implementing principles from Human-Computer Interaction (HCI), the intended output aims to significantly improve accessibility. This enhancement targets granting visually impaired users a more independent and seamless interaction within the digital realm, thereby fostering a more inclusive and accommodating web experience.

7. CONCLUSION

In summary, this project effectively addresses the pressing necessity for enhanced accessibility and usability within web applications designed for visually impaired users. Through the utilization of cutting-edge technologies such as VUI, speech recognition, assistive tools, and the integration of user-centric design principles, the project strives to narrow the accessibility gap, empowering visually impaired individuals to interact more seamlessly with digital platforms.

While each technology introduces its unique challenges, the project is committed to mitigating these obstacles through meticulous implementation, iterative design methodologies, and a keen focus on user input. Recognizing the significance of continuous enhancement and adaptation to accommodate evolving user requirements and technological progressions.

Ultimately, by amalgamating these innovative technologies and adopting a user-centric design philosophy, the project aims to foster a more inclusive digital landscape. This initiative endeavors to provide visually impaired individuals with equitable opportunities to access information, independently execute tasks, and navigate web applications with heightened ease and efficiency.

REFERENCES

[1] Artificial Intelligence-based Voice Assistant

https://ieeexplore.ieee.org/document/9210344

[2] “Hey, Siri”, “Ok, Google”, “Alexa”. Acceptance - Relevant Factors of Virtual Voice-Assistant

https://ieeexplore.ieee.org/document/8804568

[3] A Literature Review on Smart Assistant

https://www.academia.edu/download/69546476/IRJET_V8I4769.pdf

[4] DESKTOP’S VIRTUAL ASSISTANT USING PYTHON

https://www.researchgate.net/profile/Jegadeesan-Ramalingam/publication/372657833_DESKTOP'S_VIRTUAL_ASSISTANT_USING_PYTHON/links/64c22f0304d6c44bc35d350e/DESKTOPS-VIRTUAL-ASSISTANT-USING-PYTHON.pdf

[5] Intelligent personal assistants: A systematic literature review

https://www.sciencedirect.com/science/article/abs/pii/S0957417420300191