International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Speech Recognition Intelligence System for Desktop voice Assistant by using AI &IoT

Hariom Tyagi[1], VinishKumar[2], Mohammad Danish[3]Gunjan Agarwal[4], Prateek Mishra[5]

Submitted: 28/01/2023 **Accepted:** 06/04/2023

Abstract: This paper deals the Speech Recognition Intelligence System for Desktop voice Assistant by using AI &IoT, with statistical testing of hypothesis. In the modern era of reckless technology, we are able to carry out tasks that we never could have imagined we would be able to prepare for. However, in order to carry out these daydreams, we need a method that makes it simple for us to automate the things we do every day. As a result, we created applications like Voice Assistant that can communicate with us solely through human interaction. A voice assistant can be used by a number of applications, including AI and IoT. It has the ability to alter how users and machines communicate. By using voice commands, the user can access all of the features of this application, which has been designed to work with mobile phones.The primary difficulties and drawbacks of various voice assistants will be discussed in this paper. In this paper, we talk about how to make a voice-based assistant that doesn't need cloud services, which would help these devices, grow in the future.

Keywords: *SRIS, DVA, AI, IoT.*

1. Introduction

A desktop application known as a Voice Assistant enables users to perform a variety of tasks, including the language [1].

The following are our project's goals:

- Give the voice assistant a verbal command to open any website in the browser.
- Has the ability to speak the results of calculations.
- Use speech to respond to questions.
- Send an email to the desired individual.
- Launch the various desktop applications and programs that are available.
- Start YouTube and load the videos you want.
- Open Wikipedia and look up information about a particular subject.

Wouldn't it be nice for everyone to have the opulence of having a personal assistant who always attends to your calls, takes care of all of your needs, and takes the necessary actions when they arise? Voice assistants have made this opulence possible [2]. After hearing your voice command, they can do a lot of things. They can also answer questions, send emails, play music, and do other things for you [19][20][21].

Nowadays, the technology is only well-known for the user experience because these applications and services are easy to use from anywhere, regardless of where we are. Windows, Apple, Android, and so on are a few popular and well-known mobile operating systems [3]. Users of each of these Structures can take advantage of numerous applications and services. The user can connect a call or send an SMS to another user through the application. The contact information is kept in the Contacts applications. Through the Play Store, Apple Store, etc., numerous apps of a similar nature are available worldwide. All of these features result in a variety of different functionalities that can be used in mobile devices.The most significant innovations in speech recognition systems date back a very long time [4]. Recently, large-term speech recognition has been designed to work wellASR and Search in addition to Search. Voice service performance is improved through a variety of means [22][23].

It was designed to assist blind individuals who can work with their voice. It is able to recognize voices without internet. It also has mobile features like controlling a lot of applications and connecting to the network with voice commands. Additionally, it has features like Keyword

[1]*Galgotias College of Engineering & Technology, Greater Noida hariom2678@gmail.com*
[2,4]*Rajkumar Goel Institute of Technology, Ghaziabadvinishk18@gmail.com, gunjanagarwalcs@gmail.com*
[3]*Echelon Institute of Technology, Faridabad mohammaddanish@eitfaridabad.co.in*
[5]*Asia Pacific Institute of Information Technology SD India Panipat prateekmishra@apiit.edu.in*

Learning and Voice Pattern Detection. That makes using mobile devices much easier for the end user. It actively responds to the user's voice and is not dependent on the user's language. It runs faster than other applications for Online Voice Search [5].

Have a straightforward user interface that makes it easy for users to use the device [24][25].

Speech production is a non-natural method for sounding human. The goal of a speech synthesizer system is to mimic human speech.This project was funded because it would make it easier for the elderly and disabled in our society to use computers and phones. This was accomplished skillfully with the development of a speech recognition application that enables users to perform straightforward system tasks like: Using voice commands, they can check the time, open programs, search the internet for any information, and even start various applications and files on their system [6].

Using a microphone as an input device and the ASR model in Cloud Servers, the speech recognition sample will be able to recognize and transform unwanted user-generated noises into text; the text is then sent to Cloud Servers applications for analysis, and the results are provided in Fig 1 and Fig 2 as the basics of Voice Modification Module and Formation of AI &IoT voice assistant [26][27].

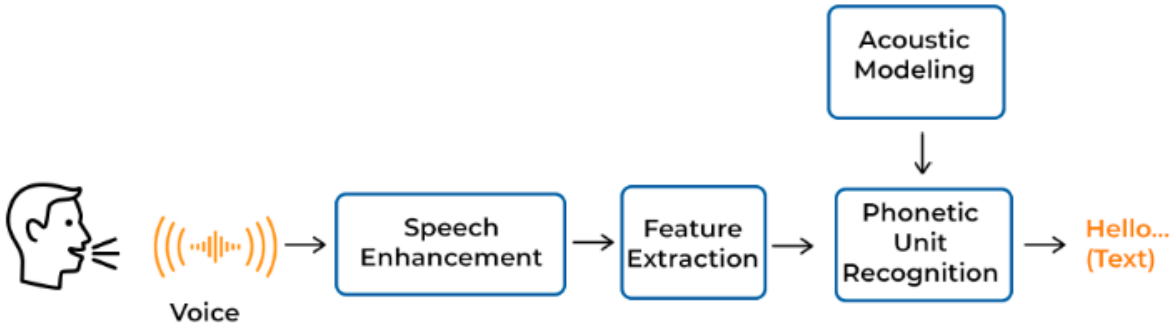

Fig 1. Voice Modification Module

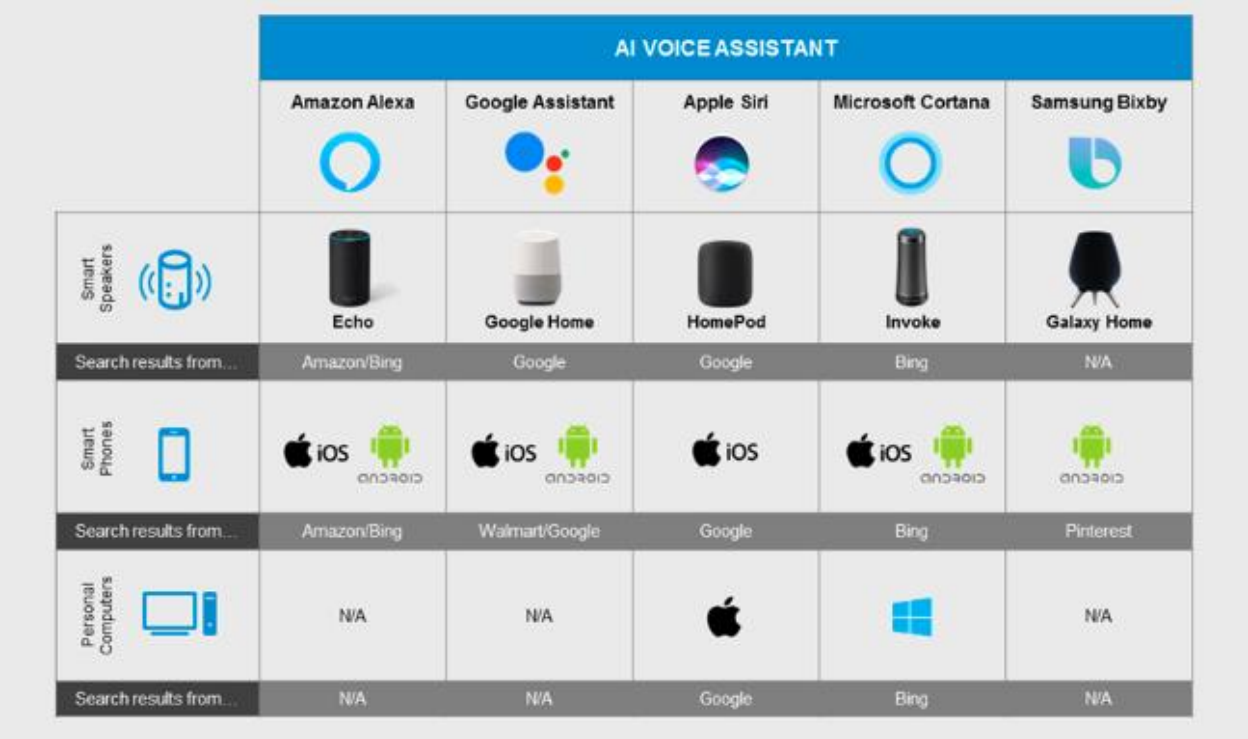

Fig 2. Formation of AI &IoTvoice assistant

2. Methodology

- Natural LanguageProcessing is the technology behind voice assistants that is utilized the most frequently. NLP relies on artificial intelligence, or intelligence performed by machines rather than humans, which necessitates a substantial data set. Automatic Speech Recognition, Natural Language Understanding, Natural Language Generation, and

Text-to-Speech are its four components [7] [28][29].

- All speech signals are converted using the ASR model into their respective words or words in a string. The vocabulary is expanding at a steady rate. A single word is easier to recognize than a continuous speech. The speaker's speech error rate can be influenced by their accent. It has been determined that the error rate of English with a Spanish or Japanese accent is three to four times higher than that of standard English [8].
- The most crucial step following ASR speech-to-text conversion is to comprehend the meaning of the text closer to the user's understanding. Ambiguity and variability are the issues with NLU [30].
- The process of putting the process into words is the process. It is the development of a language that is human-like and has some sense. The two phases of the NLG process—what to say and how to say—are further subdivided [9].
- Speech synthesis is another name for text-to-speech software. It is the final stage in making voice assistants. Text synthesis and waveform synthesis are the two steps in this process. Words are converted into speech in text synthesis, and the next step, waveform synthesis, is the creation of the desired sentence using speech samples from before in Fig 3.

Fig 3 AI&IoTvoice based system

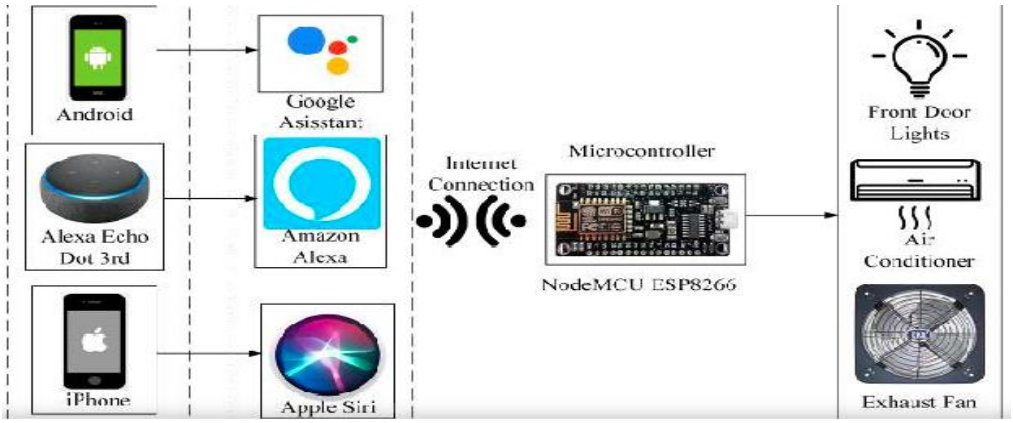

Fig.4 AI &Iot based items for voice deduction

The speech recognition module's response is parsed by Python's backend to determine in a TTS Engine to create a sound file [10]. The TTS engine is available from third-party developers in a wide range of dialects, languages, and specialist vocabulary with all AI and IoT based item are given in Fig 4.

3. Results and Discussion

A remote helper is by all accounts a speedy and proficient helper. If it is necessary, the family members or coworkers are in the systems.The four layers that make up our Desktop Voice Assistant (DVA) are as follows [11]:

- Text to speech
- Analyzing text.
- Interpret Instructions
- Speech to text

It is software that can translate audio into text. It cannot comprehend anything you might say. For computers, converted text is just letters. This text is converted into a computer command because the computer understands commands [12]. A computer command is made up of functions and their parameters. Third Layer: Internal Commands. It is doing little at this level [13].

ONE WAY ANALYSIS OF VARIANCE

One Way Analysis of Variance is used to find the significant differences between respondents' perceptions of the VRI components in relation to their various demographic profiles in Table 1.

Table 1: Analysis of the Variance among two groups from Desktop voice Assistant

Group 1 and group 2		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
AI	Between Groups	4	10	.663	2.503	.031
	Within Groups	60	250	.265		
	Total	66	255			
IoT	Between Groups	14	10	2.620	9.414	.000
	Within Groups	67	250	.278		
	Total	79	255			
SRIS	Between Groups	7	10	1.251	4.832	.000
	Within Groups	61	250	.259		
	Total	67	255			
DVA	Between Groups	9	10	1.641	3.602	.004
	Within Groups	107	250	.456		
	Total	115	255			
AI	Between Groups	0.3	10	.045	.895	.485
	Within Groups	12	250	.050		
	Total	12	255			
IoT	Between Groups	5	10	.970	5.798	.000
	Within Groups	40	250	.167		
	Total	44	255			
SRIS	Between Groups	1	10	.209	2.067	.070
	Within Groups	24	250	.101		
	Total	25	255			
DVA	Between Groups	3	10	.465	1.971	.084
	Within Groups	56	250	.236		
	Total	58	255			

Testing of Hypothesis:

The job satisfaction factor (F=2.503, P.05), the role of the family factor (F=9.414, P.05), the life satisfaction factor (F=4.832, P.05), the impact of work on family (F=3.602, P.05), the family-related factor (F=5.798, P.05), and the individual benefits factor (F=7.267, P.05) significantly differ between respondents of different ages. The work-related factor (F=.895, P>.05), personal factor (F=2.067, P>.05), social factor (F=1.971, P>.05), and psychological factor (F=2.207, P>.05) among respondents of different ages are not significantly different from one another in Fig 5.

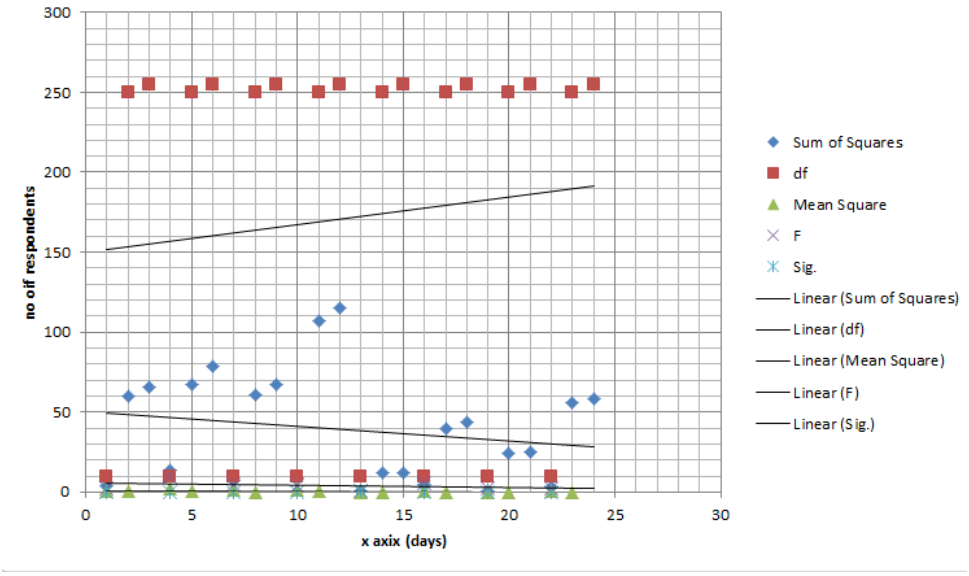

Fig 5. Statistical analysis of DVA

4. Conclusion

We discussed on Speech Recognition Intelligence System by using AI &IoT, and statistical analysis of approached for testing of hypothesison Desktop voice Assistant. To create a smart assistant, we must combine Artificial Intelligence (AI) &IoT with Natural Language Processing to create Voice Assistants. In this paper, a Python-based Voice-operated Assistant was discussed. This assistant is a program that performs fundamental tasks like updating the weather, streaming music, searching Wikipedia, and opening desktop applications. Only applications can be used with the current system's capabilities. In subsequent versions of this assistant, Artificial Intelligence (AI)&IoT will be incorporated into the system. This will result in improved recommendations based on IoT to manage nearby devices.

References

[1] Jurafsky,D.,&Martin,J.H.(2009).Chapter13:Syntactic Parsing. Speech and Language Processing:AnIntroductiontoNaturalLanguageProcessing.*ComputationalLinguistics,andSpeechRecognition(2nd ed.). Pearson–PrenticeHall*.

[2] Sen,S.,Dutta,A.,&Dey,N.(2019).SpeechProcessingandRecognitionSystem.In*AudioProcessingandSpeechRecognition*(pp.13-43).Springer,Singapore.

[3] Tomokiyo,L.M.(2001).Recognizingnon-nativespeech:characterizingandadaptingtonon-nativeusage in LVCSR. *Unpublished doctoral dissertation,CarnegieMellonUniversity,Pittsburgh,PA*.

[4] Khashabi,D.(2019).Reasoning-DrivenQuestion-Answering for Natural Language Understanding. *arXivpreprintarXiv:1908.04926*.

[5] Santhanam,S.,&Shaikh,S.(2019).ASurveyofNatural Language Generation Techniques with a FocusonDialogueSystems-Past,PresentandFutureDirections. *arXiv preprintarXiv:1906.00500*.

[6] DanielJurafsky andJamesH.Martin-Speechandlanguageprocessing,3rdeditiondraft, 2018.

[7] Srivastava S., Prakash S. (2020) Security Enhancement of IoT Based Smart Home Using Hybrid Technique. In: Bhattacharjee A., Borgohain S., Soni B., Verma G., GaoXZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science, vol 1241. Springer, Singapore. https://doi.org/10.1007/978-981-15-6318-8_44

[8] S. Srivastava and S. Prakash, "An Analysis of Various IoT Security Techniques: A Review," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 355- 362, doi: 10.1109/ICRITO48877.2020.9198027

[9] SaijshreeSrivastava, Surya Vikram Singh, RudrendraBahadur Singh, Himanshu Kumar Shukla," Digital Transformation of Healthcare: A blockchain study" International Journal of Innovative Science, Engineering &

Technology, Vol. 8 Issue 5, May 2021.

[10] V.Radha and C. Vimala, "A review on speech recognition challenges and approaches," doaj. org, vol. 2, no. 1, pp. 1–7, 2012.

[11] M. Bapat, H. Gune, and P. Bhattacharyya, "A paradigm-based finite state morphological analyzer for marathi," in Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pp. 26–34, 2010.

[12] AdityaSinha, GargiGarg, GouravRajwani, ShimonaTayal, "Intelligent Personal Assistant". International, Journal of Informative & Futuristic Research, Volune. 4, Issue 8, April 2017.

[13] Emad S. Othman ."Voice Controlled Personal Assistant Using Raspberry Pi". International Journal of Scientific and Engineering Research Volume 8, Issue 11, November-2017.

[14] Kumar, V. and Kumar, R., 2015. An adaptive approach for detection of blackhole attack in mobile ad hoc network. Procedia Computer Science, 48, pp.472-479.

[15] Kumar, V. and Kumar, R., 2015, April. Detection of phishing attack using visual cryptography in ad hoc network. In 2015 International Conference on Communications and Signal Processing (ICCSP) (pp. 1021-1025). IEEE.

[16] Kumar, V. and Kumar, R., 2015. An optimal authentication protocol using certificateless ID-based signature in MANET. In Security in Computing and Communications: Third International Symposium, SSCC 2015, Kochi, India, August 10-13, 2015. Proceedings 3 (pp. 110-121). Springer International Publishing.

[17] Kumar, Vimal, and Rakesh Kumar. "A cooperative black hole node detection and mitigation approach for MANETs." In Innovative Security Solutions for Information Technology and Communications: 8th International Conference, SECITC 2015, Bucharest, Romania, June 11-12, 2015. Revised Selected Papers 8, pp. 171-183. Springer International Publishing, 2015.

[18] Kumar, V., Shankar, M., Tripathi, A.M., Yadav, V., Rai, A.K., Khan, U. and Rahul, M., 2022. Prevention of Blackhole Attack in MANET using Certificateless Signature Scheme. Journal of Scientific & Industrial Research, 81(10), pp.1061-1072.

[19] Pentyala, S., Liu, M., & Dreyer, M. (2019). Multi-task networks with universe, group, and task feature learning. arXiv preprint arXiv:1907.01791.

[20] Srivastava, Swapnita, and P. K. Singh. "Proof of Optimality based on Greedy Algorithm for Offline Cache Replacement Algorithm." International Journal of Next-Generation Computing 13.3 (2022).

[21] Smiti, Puja, Swapnita Srivastava, and Nitin Rakesh. "Video and audio streaming issues in multimedia application." 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2018.

[22] Srivastava, Swapnita, and P. K. Singh. "HCIP: Hybrid Short Long History Table-based Cache Instruction Prefetcher." International Journal of Next-Generation Computing 13.3 (2022).

[23] Srivastava, Swapnita, and Shilpi Sharma. "Analysis of cyber related issues by implementing data mining Algorithm." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2019.

[24] P Mall and P. Singh, "Credence-Net: a semi-supervised deep learning approach for medical images," Int. J. Nanotechnol., vol. 20, 2022.

[25] Narayan, Vipul, et al. "Deep Learning Approaches for Human Gait Recognition: A Review." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.

[26] Narayan, Vipul, et al. "FuzzyNet: Medical Image Classification based on GLCM Texture Feature." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.

[27] "Keyboard invariant biometric authentication." 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT). IEEE, 2018.

[28] Mall, Pawan Kumar, et al. "Early Warning

Signs Of Parkinson's Disease Prediction Using Machine Learning Technique." Journal of Pharmaceutical Negative Results (2023): 2607-2615.

[29] Pentyala, S., Liu, M., & Dreyer, M. (2019). Multi-task networks with universe, group, and task feature learning. arXiv preprint arXiv:1907.01791.

[30] Choudhary, Shubham, et al. "Fuzzy approach-based stable energy-efficient AODV routing protocol in mobile ad hoc networks." Software Defined Networking for Ad Hoc Networks. Cham: Springer International Publishing, 2022. 125-139.