iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.41685

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Viv - The Personal Voice Assistant

Abhijeet Shetty[1], Aadit Shinde [2], Yash Sutar [3], Nidhi Sanghavi[4]
[1, 2, 3]*Student, Computer Engineering Department, Atharva College of Engineering, Mumbai University*
[4]*Assistant Professor, Computer Engineering Department, Atharva College of Engineering, Mumbai University*

Abstract: Voice control is a major growing feature that changes the way people are living their lives. Since speech recognition and natural language processing continues to advance at a rapid stage, a virtual assistant's ability and efficiency to comprehend and perform tasks will grow as well. And as voice-recognition technology improves, virtual assistant use will deeply integrate into business workflows. Tomorrow's virtual assistants are going to be built with more advanced cognitive computing technologies, which can allow a virtual assistant to know and perform multistep requests and perform more complex tasks. Virtual assistants are cloud-based programs that need internet-connected devices and/or applications to work. Examples of such services are: Siri on Apple devices, Cortana on Microsoft devices, Google Assistant for Android devices, whereas Bixby for Samsung devices.
Keywords: Desktop Assistant, Voice Recognition, Virtual Assistant, Text to Speech, Speech to Text, Language Processing, Python Script.

I. INTRODUCTION

Only in recent times for Visual Assistants can we experience major changes, the way users interact with user information. We already use them for many functions such as turning on / off the lights, playing music by streaming apps like Wynk Music, Spotify etc. This new way of communicating with technology makes lexical communication a new companion to this technology. The concept of visual aids in the past is to describe professionals who provide support services on the web. The Voice Controlled Personal Assistant System will use Natural language processing and can be integrated with artificial intelligence techniques to acquire a smart assistant that can control IoT applications. In this paper we have discussed the Personalized Voice Assistant made using python.

II. LITERATURE SURVEY

Kumaran Rangaraj, Siva, Sharan and Dhanalakshmi R [1] proposed Intelligent Personal Assistant - Implementing Voice Commands enabling Speech Recognition. Blind Source Extraction (BSE) is an approach to establish the noisy multichannel data. The implementation of Blind Source Extraction architecture necessitates and requires an extension of each system block within the framework for its flexibility and degree of blindness. The output of the enhancement algorithm amalgamated with the robust speech Recognition systems supported by gamma frequencies features, which are then analysed and on uncertainty, decoding to enhance the performance. Results are from different front-end, and back-end configurations manifest the benefits of these approaches.

Dhiraj Pratap Singh et. al [2] proposed an Voice Activated Desktop Assistant Using Python. Google's Speech Recognition API which is imported in python. This module is used to recognize the voice which is given as input by the user. The voice which is given as input is first converted to text using the speech recognition module. The text is then processed to give the result of the query given by the user. The final step is the conversion of the result of the processed query to speech which is the final output.

Subhash S et. al [3] proposed an Artificial Intelligence-based Voice Assistant. Automatic Speech Recognition which is termed as ASR is the main principle behind the working of AI-based Voice Assistant. gTTS engine package is used to make the voice assistant speak like a normal human being. gTTS is basically used to convert the audio string into text. This audio string is nothing but the response which the voice assistant is supposed to give the user. Then that audio is played using play sound package of python programming Language.

Ankush Yadav et. al [6] proposed a voice assistant for visually impaired people who wish to access, perform a relatively basic task on the internet or on the desktop itself. The basic idea being a partially, or completely visually disabled to be self-sufficient in a way that could have never been possible without the technology that we have today, voice assistants. They propose a cloud-based voice assistant – which would make up for the general concerns one has with privacy. As the data collected by the voice assistant can be used by employees and various contractors, and hence removing any worries the user might have about his/her data being misused. Furthermore, the user can send and reply to emails, messages etc. by taking input by voice. The user can also look up for any queries they might have, simply by opening up the default browser and thus searching for the answer they require. The audio signals (inputs) taken in are further translated into data that the system can understand using Google's speech to text API. This revolutionary change to convert acoustic energy to electrical energy is absolutely a need in today's advancing world.

©IJRASET: All Rights are Reserved

Giancarlo Iannizzotto et. al [7] give us an idea on how we could make a regular house into a smart home automated system. All along having a voice assistant at its core. In this paper they have combined some of the most advanced techniques in computer vision, deep learning, speech generation and recognition, and lastly artificial intelligence, into a virtual assistant architecture for smart home automation systems. The paper tells us about how the potential for voice assistants is severely bottlenecked due to the fact that it cannot use video data about and for the user & the environment. this paper introduces an architecture for building vision enabled smart assistants, provided with animated graphical characters and speech recognition.

Abhay Dekate et. al [8] wrote about the study of a voice controlled personal assistant device, its architecture, design, working and its implementation that leads us to a smarter and more innovative solutions to today's problems faced by us. It tells us about how the world is growing at a extremely rapid rate, so is technology; and voice assistants are just the beginning of a future so bright we can barely visualize it. But it is a reality after all we are interacting with a machine in almost all prospects of our lives every day. A voice assistant is meant for us to be even more efficient with our time and the rate at which we can perform certain repetitive tasks. Be it a small thing such as setting alarms or keeping you updated about the weather or the traffic in the upcoming recent/far future; it provides us a way to do things we normally would have to - only effortlessly. It applies the use of Natural language processing and can be integrated with smart artificial intelligence techniques to achieve a smart voice assistant that can control IoT applications/services and even solve user queries using web searches.

Xiangang Li et.al [12] proposed a decision tree-based state tying, in which, the states are grouped together to minimize sum-of-squared error using the DNN derived embeddings. The proposed based state tying optimizes the classification for every context dependent state, while the standard GMM based approach uses the most likelihood criterion. The main two aspects we should overcome when there is no GMM seeding: first is the producing of forced alignments. Second is the decision tree-based state tying. They used forward-backward algorithm to train DNN with unsegmented training data. For the decision tree-based state tying, the proposed approach was performed using the DNN derived embeddings, thus the GMM is not necessary. The results showed that, the proposed system outperformed the GMM based approach.

Salar Jafarlou et. al [13] shows us the importance of capturing long-term temporal dependencies of the speech signal in distant speech recognition systems. They begin by explaining the importance of the receptive field and its role in convolutional neural networks. Then comparison of conventional CNN with dilated and variants of large receptive field networks. It also explains convolutional CNNs with various receptive field size for better understanding its impact on distant speech. Finally for the result using the optimal RF size, that LRF networks performs 1.8% and 8.9% better compared to standard CNNs for clean and distant speech signals.

Yongqiang Wang et. al [15] has summarize the application of transformer and its streamable variant, Emformer based acoustic model for large scale speech recognition applications. They have compared the LSTM-based acoustic models with transformer-based ones for a range of large-scale speech recognition tasks. The results show that for low latency, voice assistant task, Emformer, a streamable transformer, gets 24% to 26% relative WERRs, compared with LSTM. For medium latency, video captioning task, compared with LCBLSTM, Emformer gets significant WERR across four languages and 2-3 times RTF reduction. Results on a task with 2.2M hours semi-supervised training data, indicate that there are still room for improvement for Emformer.

III. CONCLUSION

The system will have the following phases: Data collection in the form of voice; Voice analysis and conversion to text; Data storage and processing and generating speech from the processed text output. The data generated at every phase can further be used to find patterns and suggest users later. This can be a major base for artificial intelligence machines that learn and understand users. Therefore, on the basis of the literature review and analysis of the existing system, we have come to the conclusion that the proposed system will not only alleviate interaction with other programs and modules will also help us to keep it organized. There are still many areas to be covered in the automation world but device capabilities can help us to build a new generation of voice-controlled devices and bring continuous new change in the automation sector. This paper can serve as an example for many advanced applications.

IV. ACKNOWLEDGEMENT

We owe sincere thanks to our college Atharva College of Engineering for giving us a platform to prepare a project on the topic “Viv - The Personal Voice Assistant” and would like to thank our Principal Dr. Shrikant Kallurkar for instigating within us the need for this research and giving us the opportunities and time to conduct and present research on the topic. We are sincerely grateful for having Prof. Nidhi Sanghvi as our guide and Prof. Suvarna Pansambal, Head of Computer Engineering Department, without their motivation, constant support and valuable suggestions. Moreover, the completion of this research would have been impossible without the cooperation, suggestions and help of our friends and family.

REFERENCES

[1] Kumaran N., Rangaraj V., Siva Sharan S., Dhanalakshmi R. "Intelligent Personal Assistant - Implementing Voice Commands enabling Speech Recognition," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262279.

[2] Dhiraj Pratap Singh, Deepika Sherawat, Sonia “Voice activated desktop assistant using Python” 2020 High Technology Letter - GISTX VOL-26, 2020 doi: 10.37896/HTL26.06/115

[3] S. Subhash, P. N. Srivatsa, S. Siddesh, A. Ullas and B. Santhosh, "Artificial Intelligence-based Voice Assistant," 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020, pp. 593-596, doi: 10.1109/WorldS450073.2020.9210344.

[4] Swamy, Suma & Ramakrishnan, Kollengode. “An Efficient Speech Recognition System.” Computer Science & Engineering: An International Journal (CSEIJ), Vol. 3, No. 4. 3,2013 doi:10.5121/cseij.2013.3403.

[5] Shubham Melvin Felix, Sumer Kumar, and A. Veeramuthu "A Smart Personal AI Assistant for Visually Impaired People," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018, pp. 1245-1250, doi: 10.1109/ICOEI.2018.8553750.

[6] Ankush Yadav, Aman Singh, Aniket Sharma, Ankur Sindhu, Umang Rastogi “Desktop Voice Assistant for Visually Impaired” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9 Issue-2, July 2020 pp. 36-39, doi: 10.35940/ijrte.A2753.07922

[7] Giancarlo Iannizzotto, Lucia Lo Bello, Andrea Nucita & Giorgio Mario Grasso " A Vision and Speech Enabled, Customizable, Virtual Assistant for Smart Environments," 2018 11th International Conference on Human System Interaction (HSI), 2018, pp. 50-56, doi: 10.1109/HSI.2018.8431232.

[8] Abhay Dekate, Chaitanya Kulkarni, Rohan Killedar “Study of Voice Controlled Personal Assistant Device.” International Journal of Computer Trends and Technology. Vol. 42. 2016 pp. 42-46. doi:10.14445/22312803/IJCTT-V42P107.

[9] Veton Këpuska, Gamal Bohouta "Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018, pp. 99-103, doi: 10.1109/CCWC.2018.8301638.

[10] Mr. Yash Agarwal, Mr. Ranjeet Rai, Mr. Punit Kumar Chaubey “Brain – The A.I. (Personal voice assistant)” International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 04, 2020

[11] Tae-Kook Kim , "Short Research on Voice Control System Based on Artificial Intelligence Assistant," 2020 International Conference on Electronics, Information, and Communication (ICEIC), 2020, pp. 1-2, doi: 10.1109/ICEIC49074.2020.9051160.

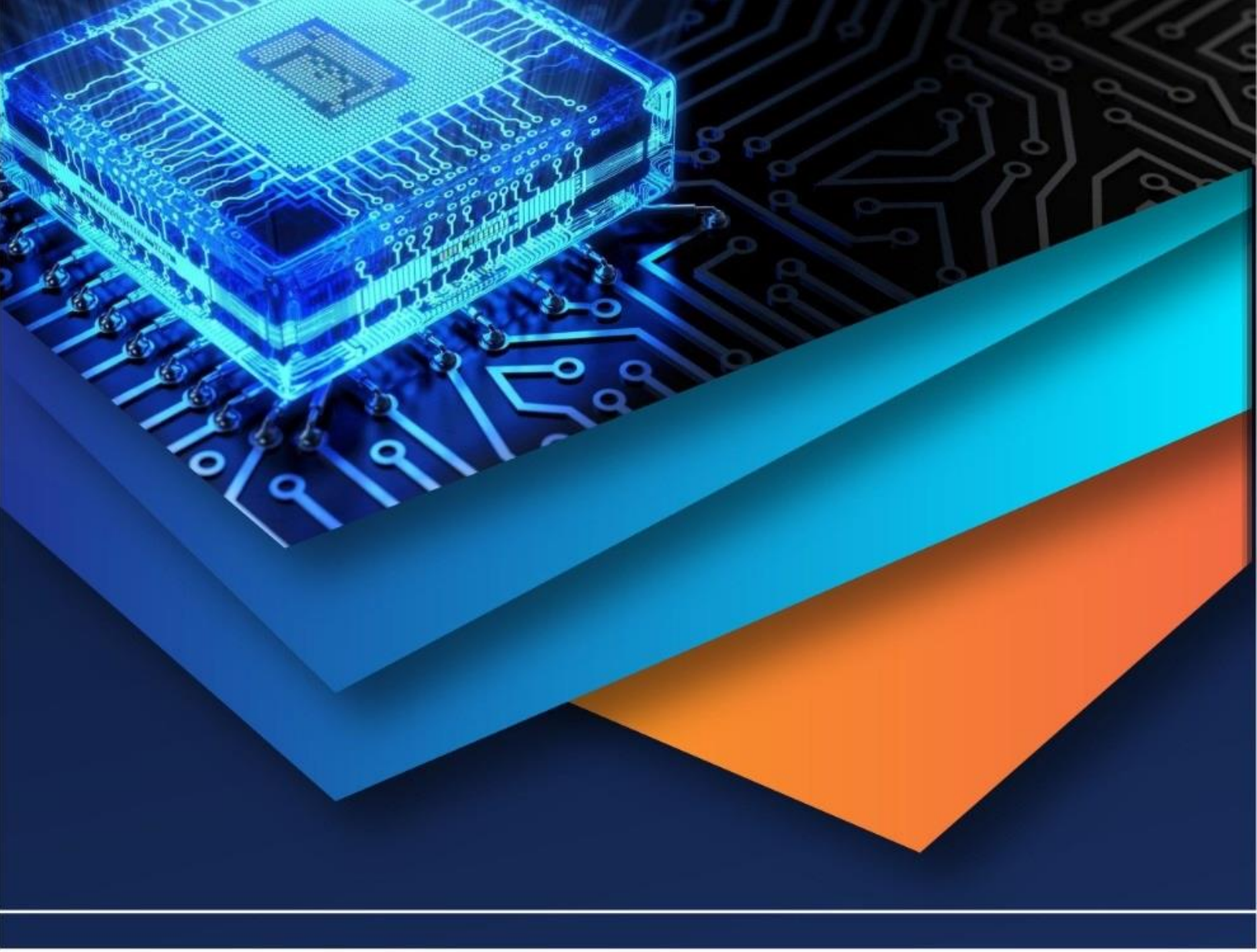
[12] Xiangang Li, Xihong Wu , "Decision tree based state tying for speech recognition using DNN derived embeddings," The 9th International Symposium on Chinese Spoken Language Processing, 2014, pp. 123-127, doi: 10.1109/ISCSLP.2014.6936637.

[13] Salar Jafarlou, Soheil Khorram, Vinay Kothapally, John H.L. Hansen "Analyzing Large Receptive Field Convolutional Networks for Distant Speech Recognition," 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2019, pp. 252-259, doi: 10.1109/ASRU46091.2019.9003805.

[14] Vikramjit Mitra, Wen Wang, Chris Bartels, Horacio Franco, Dimitra Vergyri "Articulatory Information and Multiview Features for Large Vocabulary Continuous Speech Recognition" , 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5634-5638, doi: 10.1109/ICASSP.2018.8462028.

[15] Yongqiang Wang, Yangyang Shi, Frank Zhang, Chunyang Wu, Julian Chan, Ching-Feng Yeh, Alex Xiao. "Transformer in Action: A Comparative Study of Transformer-Based Acoustic Models for Large Scale Speech Recognition Applications," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 6778-6782, doi: 10.1109/ICASSP39728.2021.9414087

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)