VIRTUAL ASSISTANT WITH HAND GESTURE INPUT

Karthik Krishnan O

Vellore Institute of Technology, Vellore

Abstract

Virtual Assistants are very common in today's world. It is basically an application program that can replace a human assistant in performing basic tasks.These assistants help the visually impaired a lot but there are rarely any of these for deaf people. The project focuses on providing a platform for the deaf people to communicate with the computer using sign languages and get their work done. This can also help ordinary people to interact with these disabled people better than before. The project is an implementation of a voice+text+hand gesture based assistant which can play songs, movies, search the web , gives quick responses to the user's questions and a lot more like other voice assistants.

Keywords: `Virtual Assistant, Chat-bot, Hand gesture recognition, Voice Assistant, Music Player , Computer Vision, Speech Recognition, Text to Speech, Wake word detection`

1. INTRODUCTION

Virtual assistants are a member of every house and office. The tasks which are historically done by personal assistants are now replaced by virtual assistants. The modern technology allows even middle class or lower class people to have a virtual assistant in their pocket. The virtual assistant systems like the google assistant or Apple's Siri, Amazon's Alexa, Microsoft Cortana are now available in one's fingertips. They are available 24/7 to listen to the users commands and perform the tasks for them.

Most of the assistants available are voice based. This turns out to be a very good aid for the visually impaired people who finds it difficult to use digital keyboards. Till date there aren't any good application software for the deaf people. They can see but cannot hear or talk. They communicate with each other using sign languages which uses their hand most of the time.

A personal assistant that can take the user's real time video as input, recognise the sign language and acts according to the user's need will be a very novel idea and also it will be a great achievement to bring up the deaf people in the society. Along with the hand gesture input, audio and text inputs are also available for the user's convenience.

Similar to other famous virtual assistants, this will also have a wake word or wake sign to run the program. The assistant will be on sleep mode for rest of the time which assures the privacy of the user.

2. LITERATURE REVIEW

S.no	Title	Findings	Tools/Concept	Reference
1	Adrisya Sahayak: A Bangla Virtual Assistant for Visually Impaired	Developing a hands free virtual assistant for native Bangla speaking people with basic functions. It supports only Bangla language and also the accuracy is not good.	Google Web Speech API, Text to speech API, Bangla vocabulary database	M. R. Sultan, M. M. Hoque, F. U. Heeya, I. Ahmed, M. R. Ferdouse and S. M. A. Mubin, "Adrisya Sahayak: A Bangla Virtual Assistant for Visually Impaired," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 597-602, doi: 10.1109/ICREST51555.2021.9331080.
2	Implementation of Virtual Assistant with Sign Language using Deep Learning and TensorFlow	An interface for recognition of sign languages of the deaf and converting them to speech. It uses alexa for their searches and returns only text. It doubles the complexity of the virtual assistant.	Tensorflow, CNN, Python Audio modules	D. Someshwar, D. Bhanushali, V. Chaudhari and S. Nadkarni, "Implementation of Virtual Assistant with Sign Language using Deep Learning and TensorFlow," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 595-600, doi: 10.1109/ICIRCA48905.2020.9183179.
3	Sign Language Alphabet Recognition Using CNN	Recognize American sign languages using Computer Vision. The accuracy in the spotting the hand is low.	MNIST classic dataset, Keras, CNN	M. Kumar, P. Gupta, R. K. Jha, A. Bhatia, K. Jha and B. K. Shah, "Sign Language Alphabet Recognition Using Convolution Neural Network," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1859-1865, doi: 10.1109/ICICCS51141.2021.9432296.
4	Gesture recognition for sign language Video Stream Translation	Sign language recognition framework based on attention mechanism	CNN LSTM Gradient weighted class activation	B. Fei, H. Jiwei, J. Xuemei and L. Ping, "Gesture recognition for sign language Video Stream Translation," 2020 5th International

3. PROPOSED WORK

Input - . After exploring different ways of implementation, the best one is used in the project. The system accepts input in text, audio and video input. It

is up-to the user to select the mode. The UI of the assistant accepts the text which can be directly processed by the algorithm. If the user wants to input in audio format, then Speech Recognition API from google is used. The audio is converted to text and then it is processed. In the case of hand sign(Video) input the CNN is used to convert that to text.

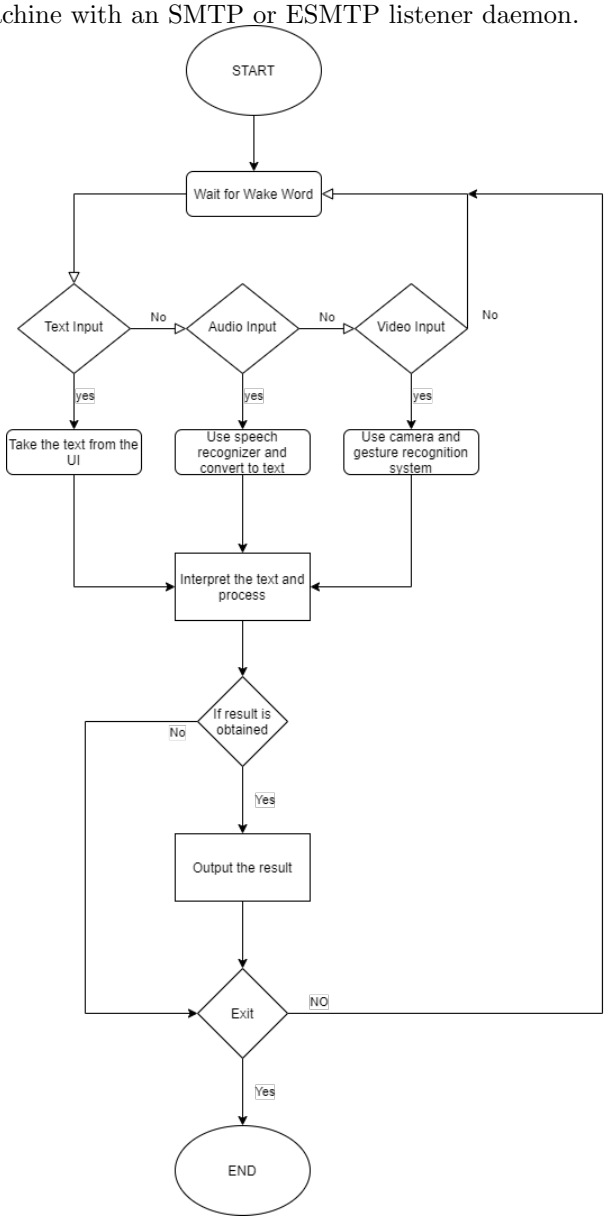
Processing. The processing of the text is done using keywords from the text. The text goes through sequence of conditional statements and the most appropriate one is selected. For example, If there a keyword "play" , the assistant will look if there is something after that. If the second word is a name of a song, that song will be played.

Information. After exploring different python APIs the one which provided the best result for out test cases is selected. For the project the Wolfram-Alpha API is used as a first priority chat-bot as well as information system. If a satisfactory result is not obtained from Wolfram , the Wikipedia API will be used. Wikipedia has a big collection of information on various topics. And if the result is still not satisfactory, the assistant will open a fresh tab in the default web browser and show what google has.

Entertainment. The biggest music bank of the present world is YouTube. From the name of the song given by the user the best match will be found from YouTube and will be played. This can also play songs for different moods, different playlists, etc. The assistant can also play music or videos which are present in the local PC.

Wake Word Detection. Almost all assistants present today has a wake word to start a competition. For Alexa, it is "hey Alexa" and for google it is "hey Google". Neural Networks are employed for acoustic modeling these days. Since our main focus is on bringing up the deaf and the blind, the need for a wake hand gesture is also required. The users image will also be fed into the assistants face recognition part for recognizing the user and greeting them whenever they come.

Messaging. The assistant can currently text in WhatsApp and Email. This will be expanded to almost all the major messaging platforms. For this purpose the inbuilt python APIs are used . Python provides smtplib module, which defines an SMTP client session object that can be used to send mail to any Internet machine with an SMTP or ESMTP listener daemon.



References