IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10539

A FRAMEWORK FOR CONTACT SAVING THROUGH VOICE RECOGNITION

Manoj Kumar S[1],Manasa G L[2],Vrathika Billava[3],Keerthi Kumar V[4], Pranav Chandran[5]

Department of Computer Science & Engineering, KSIT, Bengaluru, India[1-4]

Abstract: In modern times, everyday life has become smarter and more sophisticated. We already know some voice services like google, and Siri. etc. Now in our voice support system, it can work like automatic contact saving through voice. This project works by entering voice and rendering voice output and displaying text on the screen. Our main voice help agenda makes people smarter and deliver faster results. Voice Help captures voice input with our microphone and transforms our voice into understandable computer language providing the necessary solutions and answers that the user asks. The Natural Language Processing algorithm enables computer systems to engage in communication using the natural human language in many ways.

Keywords: Virtual Personal Assistant, Natural Human Language, Speech to text, Artificial Intelligence, Natural Language Processing, Machine Learning.

I. INTRODUCTION

Voice Recognition technology is rapidly developing, and voice search usage is also increasing due to the latest technological advancements. Nowadays almost all jobs are done digitally. We have Smartphones in our hands and nothing less than having the world in our hands. These days we don't even use our fingers to write. We are just talking about work and it is done. There are plans where we can say "Open Contacts." Text is also sent. That is the work of the Visible Assistant. It also supports specialized functions such as saving contacts and dialling and verifying it.

Voice assistants use Artificial Intelligence and Voice recognition to accurately and efficiently deliver the result that the user is looking for. While it may seem simple to ask a computer to set a timer, the technology behind it is fascinating. Contacts is the place on your mobile phone or computer where you store people's names, telephone numbers, addresses, etc. Our application is designed to save the contact details through voice.

Wise assistants based on the word need a persuasive word or a wake-up call to make the listener active, which is followed by a command. In my project the rising name is MAX. We have many visible assistants, such as Apple's Siri, Amazon's Alexa, and Microsoft's Cortana. In this project, the wake-up name is selected for MAX.

Virtual Assistants can provide several services including,

- The weather.
- Scheduling appointment time.
- Trip planning.
- Play music, movies, etc.
- Indicates the time of day.
- Manage emails. Other applications.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10539

II. RELATED WORK

According to Mackworth paper[1] proposed in 2019-20, this paper presents a comprehensive overview of the design and development of a Static Voice enabled personal assistant for pc using Python programming language. This Voice enabled personal assistant, in today's life style will be more effective in case of saving time and helpful to differently abled people, compared to that of previous days. This Assistant works properly to perform some tasks given by user. Furthermore, there are many things that this assistant is capable of doing, like sending message to user mobile, YouTube automation, gathering information from Wikipedia and Google, with just one voice command.

Through this voice assistant, we have automated various services using a single line command. It eases most of the tasks of the user like searching the web etc, We aim to make this project a complete server assistant and make it smart enough to act as a replacement for a general server administration.The project is built using open source software modules with PyCharm community backing which can accommodate any updates shortly. The modular nature of this project makes it more flexible and easy to add additional features without disturbing current system functionalities.

According to Luis Javier paper[5] Rodríguez-Fuentes, Mikel Peñagarikano, AparoVarona, Germán Bordel, "GTTS-EHU Systems for the Albayzin 2018 Search on Speech Evaluation", proceedings of Iber SPEECH, Barcelona, Spain, 2018,this paper describes the main features of KALAKA-3, a speech database specifically designed for the development and evaluation of language recognition systems. The database provides TV broadcast speech for training, and audio data extracted from YouTube videos for tuning and testing. The database was created to support the Albayzin 2012 Language Recognition Evaluation, which featured two language recognition tasks, both dealing with European languages. The first one involved six target languages (Basque, Catalan, English, Galician, Portuguese and Spanish) for which there was plenty of training data, whereas the second one involved four target languages (French, German, Greek and Italian) for which no training data was provided. Two separate sets of YouTube audio files were provided to test the performance of language recognition systems on both tasks. To allow open-set tests, these datasets included speech in 11 additional (Out-Of-Set) European languages. The paper also presents a summary of the results attained in the evaluation, along with the performance of state-of-the-art systems on the four evaluation tracks defined on the database, which demonstrates the extreme difficulty of some of them. As far as we know, this is the first database specifically designed to benchmark spoken language recognition technology on YouTube audios.

The research paper[11] of Y. Liu and K. Kirchhoff, ''Graph-based semisupervised learning for acoustic modeling in automatic speech recognition,'' IEEE/ACM Trans. Audio, Speech, Language Process., In this paper, we investigate how to apply graph-based semisupervised learning to acoustic modeling in speech recognition. Graph-based semisupervised learning is a widely used transductive semisupervised learning method in which labelled and unlabelled data are jointly represented as a weighted graph; the resulting graph structure is then used as a constraint during the classification of unlabelled data points. We investigate suitable graph-based learning algorithms for speech data and evaluate two different frameworks for integrating graph-based learning into state-of-the-art, deep neural network DDN-based speech recognition systems. The first framework utilizes graph based learning in parallel with a DNN classifier within a lattice-rescoring framework, whereas the second framework relies on an embedding of graph neighborhood information into continuous space using an autoencoder. We demonstrate significant improvements in frame level phonetic classification accuracy and consistent reductions in word error rate on large vocabulary conversational speech recognition tasks.

III. PROBLEM IDENTIFICATION

PROBLEM STATEMENT: The present system like Apple Siri, Google Assistant can call the existing contacts but does not accept the command "Create a contact" or "Add a new contact" and also does not give any confirmation to the user about the contact details.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10539

OUR SOLUTION TO THE PROBLEM: The application which we are designing overcomes all these troubles and it will be very easy to handle the contacts.

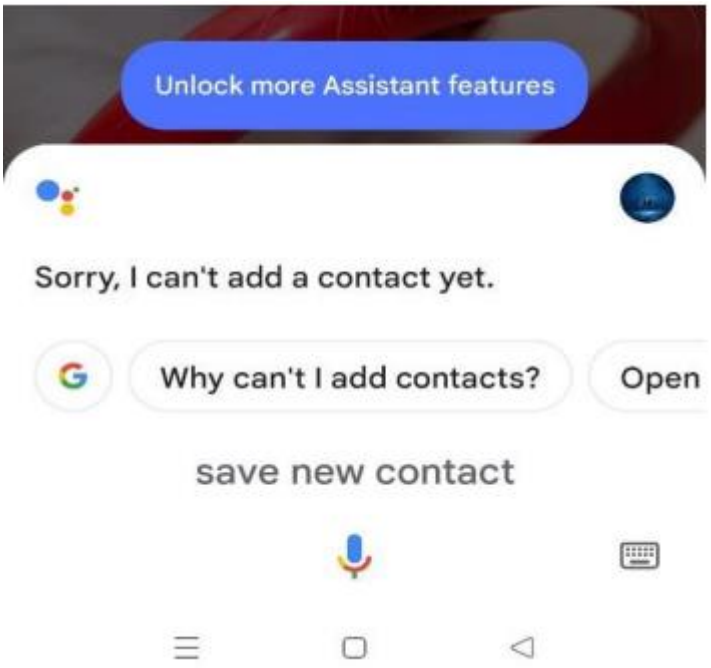

Fig. 1 **Google Assistant is unable to save new contact details through voice command.**

IV. METHODOLOGY

A. PROPOSED SYSTEM

Speech Recognition module: The system uses Google's online speech recognition system for converting speech input to text. The speech input Users can obtain texts from the special corpora organized on the computer network server at the information center from the microphone is temporarily stored in the system and is then sent to Google cloud for speech recognition. The equivalent text is then received and fed to the central processor.

Python/Java Backend: The backend gets the output from the speech recognition module and then identifies whether the command or the speech output is an API Call and Context Extraction. The output is then sent back to the backend to give the required output to the user.

API calls: API stands for Application Programming Interface intermediary that allows two applications to talk to each other. In other words, an API is a messenger that delivers your request to the provider that you're requesting it from and then delivers the response back to you.

Content Extraction: Context extraction (CE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine readable documents. In most cases, this activity concerns processing human language texts using natural language processing (NLP).

Recent activities in multimedia document processing like automatic annotation and content extraction out of images/audio/video could be seen as context extraction TEST RESULTS.

Text-to-speech module: Text-to-Speech (TTS) refers to the ability of computers to read text aloud. A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10539

output as sound. TTS engines with different languages, dialects and specialized vocabularies are available through thirdparty publishers.

B. DATA FLOW DIAGRAM

A dataflow diagram is a way of representing a flow of data through a process or a system. By using a simple graphical representation it can be transformed as the inputs for the system.

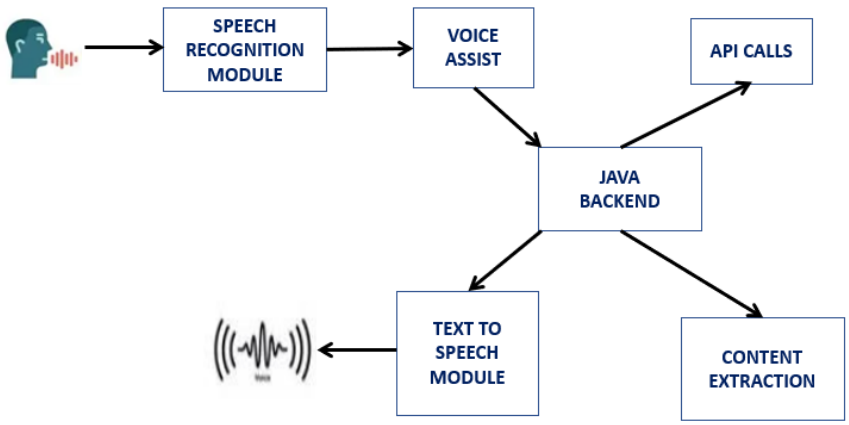

Fig. 2 Dataflow Diagram

V. SNAPSHOTS

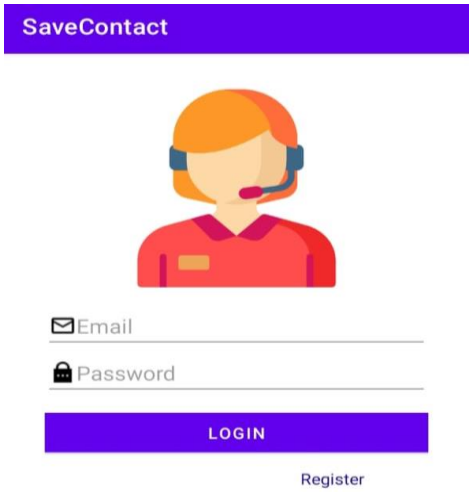

Fig. 3 Login Credentials

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10539

Fig. 4. Enter Required Details

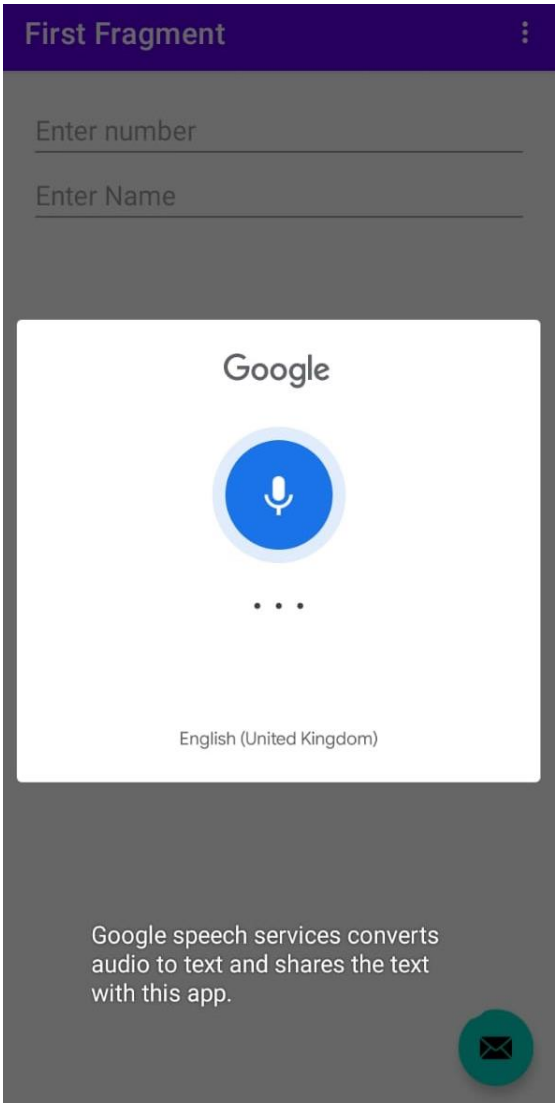

Fig. 5. Speech Recognition

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10539

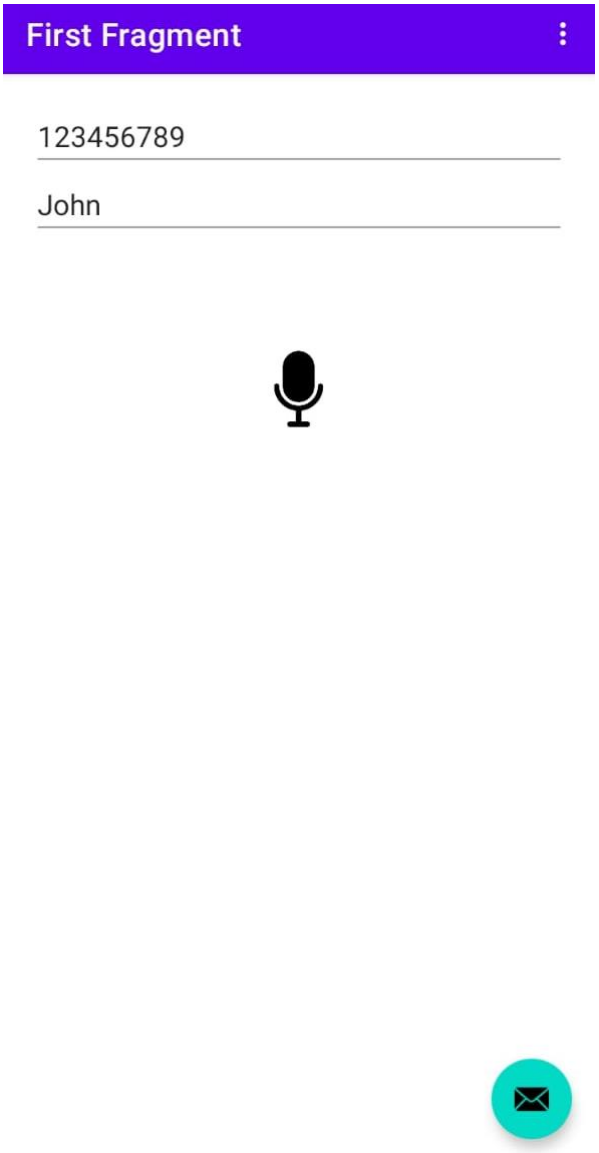

Fig. 6. Contact Details is saved through voice

CONCLUSION

In this paper "A framework for saving contacts through voice recognition" we discussed the design and implementation of Digital Assistance. The project is built using open source software modules with PyCharm community backing which can accommodate any updates shortly. The modular nature of this project makes it more flexible and easy to add additional features without disturbing current system functionalities. It not only works on human commands but also give responses to the user based on the query being asked or the words spoken by the user such as opening tasks and operations. It is saving the contact details the way the user feels more comfortable and feels free to interact with the voice assistant. The application should also eliminate any kind of unnecessary manual work required in the user life of performing every task. The entire system works on the verbal input rather than the next one.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10539

REFERENCES

[1] Mackworth (2019-2020), Python code for voice assistant: Foundations of Computational Agents- David L. Poole and Alan K. Mackworth.

[2] Nil Goksel, CanbekMehmet ,EminMutlu, "On the track of Artificial Intelligence: Learning with Intelligent Personal Assistant", proceedings of International Journal of Human Sciences, 2018.

[3] Luis Javier Rodríguez-Fuentes, Mikel Peñagarikano, AparoVarona, Germán Bordel, "GTTS-EHU Systems for the Albayzin 2018 Search on Speech Evaluation", proceedings of IberSPEECH, Barcelona, Spain, 2018.

[4] Ravivanshikumar ,Sangpal,Tanvee ,Gawand,SahilVaykar, "JARVIS: An interpretation of AIML with integration of gTTS and Python", proceedings of the 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kanpur, 2019.

[5] Luis Javier Rodríguez-Fuentes, Mikel Peñagarikano, AparoVarona, Germán Bordel, "GTTS-EHU Systems for the Albayzin 2018 Search on Speech Evaluation", proceedings of IberSPEECH, Barcelona, Spain, 2018

[6] T. Kinnunen and H. Li, ''An overview of text independent speaker recognition: From features to supervectors,'' Speech Commun., vol. 52, no. 1, pp. 12– 40, 2010, doi: 10.1016/j.specom.2009.08.009.

[7] M. S. Gazzaniga, ''Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition?'' Brain, vol. 123, no. 7, pp. 1293–1326, Jul. 2000, doi: 10.1093/brain/123.7.1293.

[8] L. Parra and C. Spence, ''Convolutive blind separation of non-stationary sources,'' IEEE Trans. Speech Audio Process., vol. 8, no. 3, pp. 320–327, May 2000, doi: 10.1109/89.841214.

[9] G. Pironkov, S. U. Wood, and S. Dupont, ''Hybrid-task learning for robust automatic speech recognition,'' Comput. Speech Lang., vol. 64, Nov. 2020, Art. no. 101103, doi: 10.1016/j.csl.2020.101103.

[10] W. Li, P. Zhang, and Y. Yan, ''TEnet: Target speaker extraction network with accumulated speaker embedding for automatic speech recognition,'' Electron. Lett., vol. 55, no. 14, pp. 816–819, Jul. 2019, doi: 10.1049/el.2019.1228.

[11] Y. Liu and K. Kirchhoff, ''Graph-based semisupervised learning for acoustic modeling in automatic speech recognition,'' IEEE/ACM Trans. Audio, Speech, Language Process., vol. 24, no. 11, pp. 1946–1956, Nov. 2016, doi: 10.1109/TASLP.2016.2593800