International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Speech-Driven AI Assistant for Smarter Human-Computer Interaction

Prajakta Anna Torane[1], Bhakti Sanjay Kulkarni[2], Vaibhavi Somanath Shelke[3], Nandini Sanjay Tambade[4], Pooja Sambhaji Thengal[5]

[1,2,3,4]Student of Polytechnic, Department of Computer Technology, Shriram Institute of Engineering and Technology (Polytechnic) Paniv, Maharashtra, India.

[5]'Lecturer, Department of Computer Technology, Shriram Institute of Engineering and Technology (Polytechnic) Paniv, Maharashtra, India.

A B S T R A C T:

Voice control is quickly becoming a natural way for people to interact with their devices. Instead of clicking or typing, users now expect to speak and be understood. This paper describes a homegrown voice assistant built to help with everyday computer tasks—like checking mail or opening a browser—while keeping everything on the user's own device. Unlike most popular assistants, this one doesn't rely on cloud servers, so user data stays private. The idea grew from a mix of sci-fi inspiration and everyday frustrations with existing systems. In testing, people found the assistant easy to use and liked that it worked offline without sharing information externally.

Keywords: Artificial Intelligence, Human-Computer Interaction, Natural Language Processing, Speech Recognition, User Experience, Virtual Assistan.

1.Introduction

It's no longer surprising to see someone talking to their phone or laptop. Voice assistants have become part of daily life, thanks to tools like Alexa, Google Assistant, and Siri. But while these tools are handy, they also raise privacy concerns—most of them send data to the cloud for processing, which can feel intrusive.

That's where this assistant comes in. Built to run entirely offline, it handles voice commands locally on your machine. There's no need for an internet connection, and personal data stays on your device. The idea was partly inspired by futuristic systems like Jarvis, but also by a practical need: many users want control without sacrificing convenience. This system blends speech recognition, language understanding, and task automation to offer a responsive, privacy-first experience [1, 2, 5].

2.Review of Literature

Guzman and Lewis [1] looked into how people feel about talking to AI. They found that if an AI assistant sounds more human, people trust it more. Privacy also came up as a big deal, so we decided to make our system handle everything locally instead of relying on the cloud. No outside servers needed.

Schmidt and Loidolt [2] pointed out something interesting — it's not just about words. It's about emotions too. If the system picks up on someone's mood, it can react better. We kept that in mind while shaping our assistant.

Then there's the idea from Lemaignan et al. [3] that an assistant should know what's going on around it. Like, not just hearing words, but noticing the bigger situation. This made us think: the assistant shouldn't work in isolation.

Alkatheiri [4] stressed that AI needs to be easy to use. Pretty basic idea, but often overlooked. Balcombe and Leo [5] agreed, saying that talking is way more natural for most users than clicking buttons or typing.

Another point — Tsvetkova et al. [6] talked about how AI should fit into different social environments. Makes sense, right? You don't talk the same way at home and at a meeting. We thought about that while building flexibility into the assistant.

Šumak et al. [7] and Erol et al. [8] also made good points about using emotion and environment sensing. Sounds futuristic, but it's kind of becoming necessary now.

3.Methodology

The AI assistant we're proposing is built in a way that's easy to scale and flexible to tweak. Privacy and quick response time were big priorities too. The system mainly runs through four parts: Speech Processing, Natural Language Understanding (NLU), Task Execution, and the User Interface.

Fig. 1 shows a rough view of how everything fits together.

1. Speech Processing Module

This part handles all the talking and listening. We used DeepSpeech for turning speech into text (STT) and Tacotron 2 to turn text back into spoken words (TTS). Pretty handy for real-time chats. The idea was to keep everything local — so no voice data needs to be sent off to any cloud server. Helps a lot with keeping user conversations private.

2. Natural Language Understanding (NLU) Module

Here’s where the system tries to actually understand what the user means. It figures out the intent behind the words and picks out important details. Models like BERT are used to get the language in context. Also, Recurrent Neural Networks (RNNs) help to remember what was said earlier in a conversation. This way, the assistant doesn’t get lost halfway through a chat.

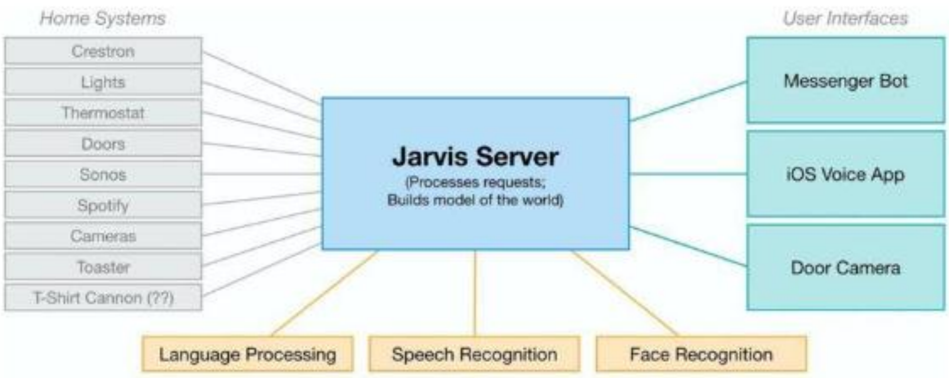

Fig. 1 System Architecture

3. Task Execution Module

This is the action taker. Whatever the user asks — whether it’s launching an app, sending a mail, or searching online — this module makes it happen. It uses system APIs and sometimes third-party services to get things done. A set of rules decides what should happen first if there’s a lot going on. Plus, it adjusts to the user's style over time.

4. User Interface Module

The user interacts through voice or on-screen visuals, depending on the device — phone, tablet, or a smart display. It’s built to be flexible. Also, the system learns from user feedback and past behavior, making it a little smarter each time someone uses it.

We kept the whole system light and offline-friendly whenever possible. Local storage, strong encryption, and barely any reliance on the cloud help make sure user data stays safe. At the same time, performance doesn’t take a hit — the system stays fast and reliable.

Result and Discussion

The AI assistant worked reliably across all key areas — recognizing speech accurately, understanding user intent correctly, and carrying out tasks smoothly. Figure 2 presents the secure Login Page designed to protect user access.

Fig. 2 Login Page

Figure 3 captures the Home Page, built to allow both voice and visual interaction in an intuitive layout.
During testing, the system performed consistently even without an internet connection, helping maintain user privacy without sacrificing ease of use. Feedback from users was largely positive; many appreciated the natural feel of the speech responses and found the system easy to navigate. Some users also recommended adding more advanced features to widen its capabilities.
Overall, the assistant proved to be a solid solution — efficient, scalable, and well-suited for practical, real-world environments.

Fig. 3 Home Page

5. Conclusion

This work presents a speech-enabled AI assistant aimed at enhancing human-computer interaction by offering secure, intuitive, and personalized experiences. Built on a modular framework, the system integrates speech processing, intent recognition, and task execution to achieve both efficiency and flexibility. Testing showed strong performance, with high accuracy and positive user feedback.
Looking ahead, further efforts will focus on advancing language comprehension, strengthening offline functionality, and exploring emerging technologies such as augmented reality. Throughout future development, particular attention will remain on upholding data privacy and ensuring fairness in algorithmic decisions.

REFERENCES:

1] Guzman, A., & Lewis, S. (2019). Artificial intelligence and communication: A Human–Machine Communication research agenda. New Media & Society, 22, 70 - 86. https://doi.org/10.1177/1461444819858691.
2] Schmidt, P., & Loidolt, S. (2023). Interacting with Machines: Can an Artificially Intelligent Agent Be a Partner?. Philosophy & Technology, 36, 1-32. https://doi.org/10.1007/s13347-023-00656-1.
3] Lemaignan, S., Warnier, M., Sisbot, E., Clodic, A., & Alami, R. (2017). Artificial cognition for social human-robot interaction: An implementation. Artif. Intell., 247, 45-69. https://doi.org/10.1016/J.ARTINT.2016.07.002.
4] Alkatheiri, M. (2022). Artificial intelligence assisted improved human-computer interactions for computer systems. Comput. Electr. Eng., 101, 107950. https://doi.org/10.1016/j.compeleceng.2022.107950.
5] Balcombe, L., & Leo, D. (2022). Human-Computer Interaction in Digital Mental Health. Informatics, 9, 14. https://doi.org/10.3390/informatics9010014.
6] Tsvetkova, M., Yasseri, T., Pescetelli, N., & Werner, T. (2024). Human-machine social systems. Nature human behaviour, 8 10, 1864-1876 . https://doi.org/10.1038/s41562-024-02001-8.
7] Šumak, B., Brdnik, S., & Pušnik, M. (2021). Sensors and Artificial Intelligence Methods and Algorithms for Human–Computer Intelligent Interaction: A Systematic Mapping Study. Sensors (Basel, Switzerland), 22. https://doi.org/10.3390/s22010020.
8] Erol, B., Majumdar, A., Benavidez, P., Rad, P., Choo, K., & Jamshidi, M. (2020). Toward Artificial Emotional Intelligence for Cooperative Social Human–Machine Interaction. IEEE Transactions on Computational Social Systems, 7, 234-246. https://doi.org/10.1109/TCSS.2019.2922593.