**CHAPTER 4**

**SYSTEM DESIGN AND DEVELOPMENT**

1. **User Interaction Layer**:

* **Input**: Voice (microphone), Text (chat interface).
* **Output:** Speech (speaker), Text (display).

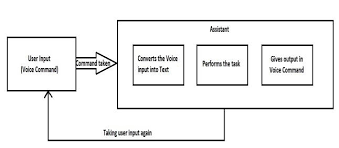
1. **Processing Layer**:

* **Speech Recognition**: Converts spoken language into text.
* **Natural Language Understanding (NLU):** Interprets the meaning of the text.
* **Dialogue Management**: Manages the flow of conversation.
* **Natural Language Generation (NLG):** Generates the response text.
* **Text-to-Speech (TTS):** Converts response text back into speech.

**3. Text-to-Speech (TTS)**:

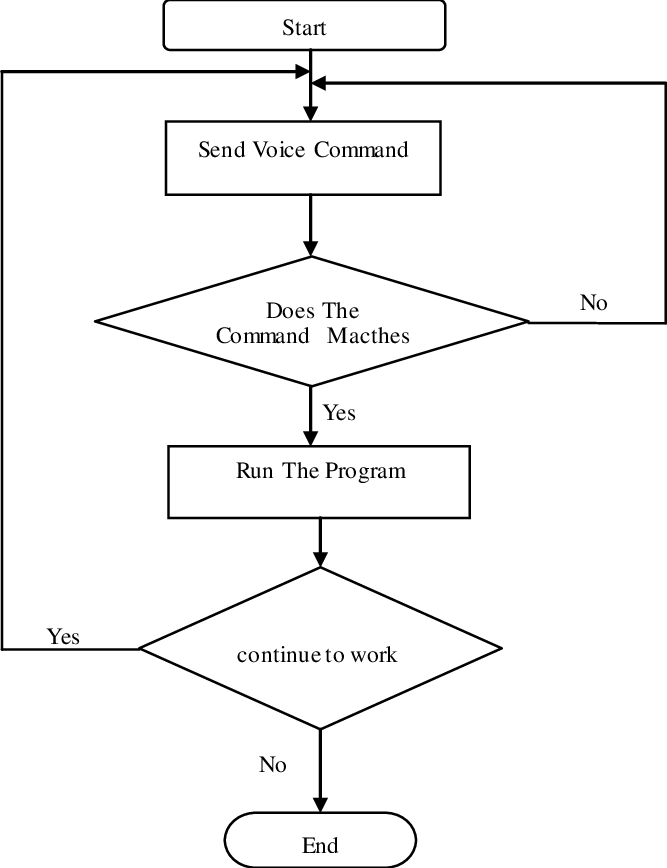
* Google Cloud Text-to-Speech

**4.1 Architectural Design**



.

This Assistant consists of three modules. First is, assistant accepting voice input from user. Secondly, analysing the input given by the user, and mapping it to the respective intent and function. And the third is, the assistant giving user the result all along with voice. Initially, the assistant will start accepting the user input. After receiving the input, the assistant will convert the analog voice input to the digital text. If assistant was not able to convert the voice into text, it will start asking user for the input again. If converted, it will start analyzing the input and will map the input with particular function. And later, the output will be given to user via the voice command



The flow of a voice assistant starts with the user initiating interaction, capturing voice input via a microphone. This audio is converted to text using a speech recognition module. The text is then processed by the Natural Language Understanding (NLU) module to identify the user’s intent and extract relevant entities. The dialogue management system uses this information to determine the appropriate action, often accessing backend services to fetch data or perform tasks. The response is generated in text form by the Natural Language Generation (NLG) module and then converted back to speech using Text-to-Speech (TTS). Finally, the spoken response is delivered to the user through the speaker, completing the interaction.