Project Report

Project_Title Speech Recognition and Speech Synthesis System

Subject Deep Learning

Semester BS(DS) 6th

Submitted to Dr.Waqar Malik

Submitted By

Name	Roll no
Hadia Habib	(12)

Submitted Date 17/12/2025

Session 2022-2026

DEPARTMENT OF DATA SCIENCE

FACULTY OF COMPUTING AND ENGINEERING

Contents

Abstract

This project focuses on developing a personalized voice assistant using Python that can understand spoken input and respond through text and speech. The system uses speech recognition for voice-to-text conversion and text-to-speech synthesis with male and female voice options. A predefined knowledge base is used to answer common queries, and a Gradio-based interface enables easy interaction. The assistant is designed mainly for educational and personal use.

Keywords

Voice Assistant, Speech Recognition, Text-to-Speech, Python, Gradio, Human–Computer Interaction

1. Introduction

Voice assistants are software agents that can understand spoken language, process it, and respond using natural language. They are increasingly used in smart devices, healthcare, education, and personal productivity. Popular examples include Amazon Alexa, Google Assistant, and Apple Siri.

This project aims to develop a **personalized voice assistant** capable of:

- Understanding spoken queries (voice → text).
- Responding intelligently using pre-defined knowledge (text → text).
- Generating spoken responses (text → voice) with selectable male/female voice options.

Unlike general-purpose assistants, this project is tailored for **educational and personal usage**, using Python libraries for speech recognition and text-to-speech synthesis.

3. Problem Statement

Many users require simple, offline-capable assistants for learning, information retrieval, or personal help. Existing solutions may be complex, require cloud access, or lack flexibility.

This project addresses these problems by:

- Allowing **Speech-based interaction** without needing extensive setup.
- Providing **textual and auditory feedback**, enhancing accessibility.
- Using a **predefined knowledge base** for consistent responses.

4. Objectives

1. Build a **voice-based AI assistant** using Python.
2. Enable **speech recognition** to convert voice to text using `SpeechRecognition`.
3. Implement **text-to-speech** responses using `gTTS` and `pydub`.
4. Provide **male/female voice options** for responses.
5. Create an **interactive GUI** with Gradio for voice and text inputs.
6. Test the assistant on common queries (time, AI, Python, health advice, etc.).

5. Scope of the Project

- **Input:** Voice recordings or typed text.
- **Output:** Textual reply and synthesized voice reply.
- **Knowledge Base:** Predefined set of 50+ common queries and answers.
- **Technologies Used:** Python, Gradio, SpeechRecognition, gTTS, pydub.
- **Limitations:**
    - Does not fetch live data (weather/news) online.

- Cannot perform complex tasks like opening apps or controlling devices.

6. Methodology

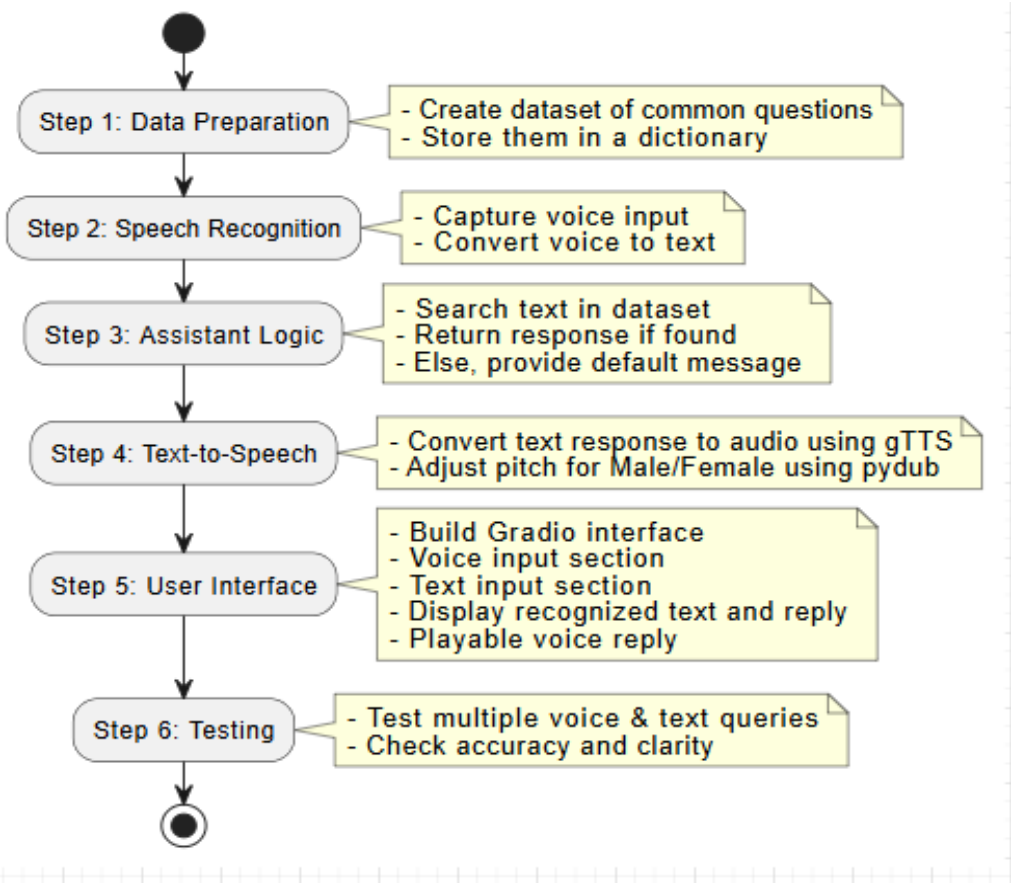

FIGURE 1 METHODOLOGY OF SPEECH RECOGNITION & SPEECH *SYNTHESIS SYSTEM*

The following activity diagram illustrates the step-by-step workflow of the voice assistant project, from data preparation to testing, including voice recognition, assistant logic, and text-to-speech response generation.

7. Tools and Technologies

Tool/Library	Purpose
Python	Programming language
SpeechRecognition	Convert voice to text
gTTS	Text-to-speech conversion
pydub	Audio manipulation (male/female voice)
Gradio	User-friendly GUI for voice/text interaction
datetime	Handle time-related queries

8. Expected Outcomes

- A fully functional **voice assistant** that responds in text and speech.
- Interactive interface where users can ask questions **via voice or text**.
- Adjustable voice gender option for personalized interaction.
- Dataset-based knowledge for handling 50+ common queries.

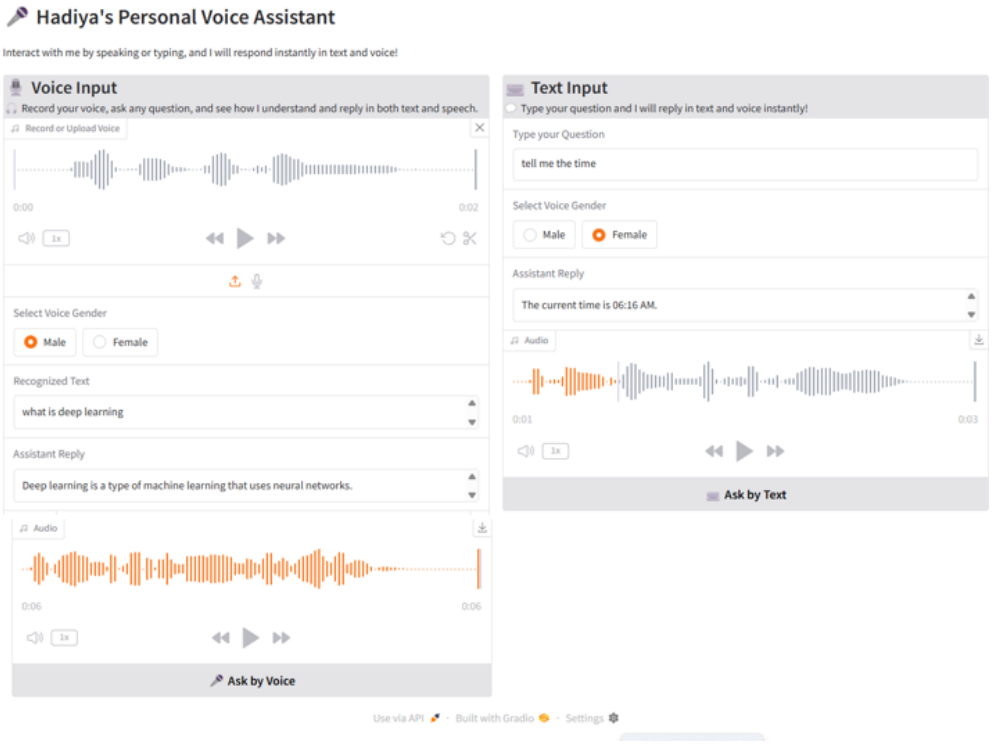

FIGURE 2 OUTPUT

9. Applications

- **Educational Tool:** Helps students learn AI concepts or programming.
- **Personal Assistant:** Can remind or answer basic questions.
- **Accessibility:** Aids visually impaired users who rely on voice interfaces.

10. Timeline

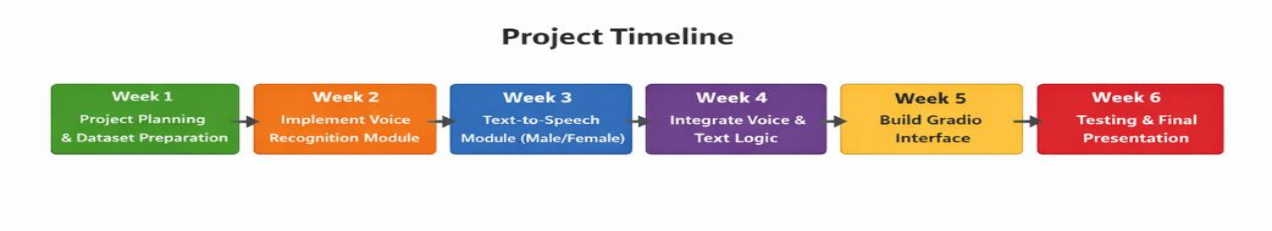

FIGURE 3 TIMELINE

11. References

1. SpeechRecognition Documentation: https://pypi.org/project/SpeechRecognition/
2. gTTS Documentation: https://gtts.readthedocs.io/
3. pydub Documentation: https://pydub.com/
4. Gradio Documentation: https://gradio.app/