AI Voice Assistant Bot
With OpenAI TTS

Zhiyu (Zoey) Zhang

Project Design

- **API Integration:**
  - **GPT-4o-mini** for text generation.
  - **Whisper API** for speech-to-text transcription.
  - **OpenAI TTS** for text-to-speech.

- Supports **threaded** and **linear** modes.

- Designed for **conversational flow**, **high-quality audio playback**, and **graceful shutdown and restarts**.

Key Features

- **AI Models:**
  - **GPT-4o-mini** balances cost and performance.
  - **Whisper API** for accurate transcription.
  - **OpenAI TTS** replaces gTTS for professional audio quality.

- **Functionality:**
  - Clear queues during shutdown using `clear_queue()`.
  - **Verbose mode**: Debugging logs.
  - **Default mode**: Clean question-and-answer output.

Key Features

- **Modes:**

  - **Threaded mode**: Real-time interaction.

  - **Linear mode**: Sequential processing for cleaner audio.

- **Improvements over GTTS Project:**

  - Integration of **OpenAI TTS** for **richer audio playback**.

  - Enhanced graceful shutdown with **resource management**.

Core Functionalities

- **record_audio()**: Captures and queues audio input.
- **transcribe_forever()**: Converts audio to text using Whisper API.
- **reply()**: Generates GPT responses and outputs speech using OpenAI TTS.
- **clear_queue()**: Releases queue resources during shutdown.
- **get_completion()**: Handles GPT-4o-mini interactions with the latest message structure.

Sample Use Cases

Threaded Mode with verbose:

- Detailed logs with timestamps.

- Smooth transitions between conversations with wake and stop words.

- Ignore input when not in conversation mode

```
(base) zhangzhiyu@shengchanligongju openai % python sts_openai_threads.py --verbose
Listening...
[2024-11-24 19:19:20.962496] Transcription result: 'Hey computer, what's the color of the sky?
'
[2024-11-24 19:19:20.962647] Wake word detected. Starting conversation mode.
[2024-11-24 19:19:20.962733] Processing input: ' whats the color of the sky'
[2024-11-24 19:19:21.805038] User question: ' whats the color of the sky'
[2024-11-24 19:19:21.805067] Bot response: 'The color of the sky is typically blue during the
day due to the scattering of sunlight by the Earth's atmosphere. However, it can also appear i
n various shades, such as gray during overcast conditions, orange or pink during sunrise and s
unset, and black'
[2024-11-24 19:19:40.720797] Playback completed.
[2024-11-24 19:19:40.721196] Temporary audio file 'reply.mp3' removed.
[2024-11-24 19:19:47.276939] Transcription result: 'Thank you, stop.'
[2024-11-24 19:19:47.276985] Stop word detected. Ending conversation.
[2024-11-24 19:19:52.907571] Transcription result: 'What's the color of grass?'
[2024-11-24 19:19:52.907602] Ignoring input (not in conversation mode).
[2024-11-24 19:20:05.093854] Transcription result: 'Hey computer, what's the date today?'
[2024-11-24 19:20:05.093918] Wake word detected. Starting conversation mode.
[2024-11-24 19:20:05.093935] Processing input: ' whats the date today'
[2024-11-24 19:20:05.708048] User question: ' whats the date today'
[2024-11-24 19:20:05.708079] Bot response: 'Today's date is October 4, 2023.'
[2024-11-24 19:20:10.024357] Playback completed.
[2024-11-24 19:20:10.024749] Temporary audio file 'reply.mp3' removed.
[2024-11-24 19:20:14.618642] Transcription result: 'Thank you, stop.'
[2024-11-24 19:20:14.618688] Stop word detected. Ending conversation.
^C
Shutting down...
All threads terminated. Goodbye!
```

Sample Use Cases

Threaded Mode by default:

```
[(base) zhangzhiyu@shengchanligongju openai % python sts_openai_threads.py
Listening...
User question: ' whats the color of the sky'
Bot response: 'The color of the sky is typically blue during the day due to the scattering of
sunlight by the Earth's atmosphere. However, it can also appear in various shades such as gray
, orange, pink, or red during sunrise and sunset, and can be over'

User question: 'Whats the color of grass'
Bot response: 'The color of grass is typically green. However, it can vary in shade depending
on factors such as the type of grass, the season, and environmental conditions.'

^C
Shutting down...
All threads terminated. Goodbye!
```

Sample Use Cases

Linear Mode with verbose:

- Clearer playback without feedback.

- Sequential task processing.

```
[(base) zhangzhiyu@shengchanligongju openai % python sts_openai_linear.py --verbose
Listening...
[2024-11-24 19:54:56.669462] Waiting for input...
[2024-11-24 19:55:00.640389] Captured audio.
[2024-11-24 19:55:02.537331] Wake word detected. Starting conversation mode.
[2024-11-24 19:55:02.537383] User input: ', what's the color of the sky?'
[2024-11-24 19:55:03.124620] Bot response: 'The color of the sky is typically blue during the
day due to the scattering of sunlight by the Earth's atmosphere. However, it can also appear i
n various colors at different times, such as orange or pink during sunrise and sunset, and gra
y when it's over'
[2024-11-24 19:55:22.969952] Playback completed.
[2024-11-24 19:55:22.970329] Temporary audio file 'reply.mp3' removed.
[2024-11-24 19:55:22.970387] Waiting for input...
[2024-11-24 19:55:27.090021] Captured audio.
[2024-11-24 19:55:28.407255] Stop word detected. Ending conversation.
[2024-11-24 19:55:28.407310] Waiting for input...
[2024-11-24 19:55:33.615517] Captured audio.
[2024-11-24 19:55:34.881859] Ignoring input (not in conversation mode).
[2024-11-24 19:55:34.881896] Waiting for input...
[2024-11-24 19:55:46.291919] Captured audio.
[2024-11-24 19:55:47.374445] Wake word detected. Starting conversation mode.
[2024-11-24 19:55:47.374468] User input: ', what's the date today?'
[2024-11-24 19:55:47.886551] Bot response: 'Today's date is October 4, 2023.'
[2024-11-24 19:55:52.451366] Playback completed.
[2024-11-24 19:55:52.451684] Temporary audio file 'reply.mp3' removed.
[2024-11-24 19:55:52.451746] Waiting for input...
^C
Exiting...
```

Sample Use Cases

Linear Mode by default:

```
[(base) zhangzhiyu@shengchanligongju openai % python sts_openai_linear.py
Listening...
User question: ', what's the color of the sky?'
Bot response: 'The color of the sky is typically blue during the day due to the scattering of
sunlight by the atmosphere. However, it can also appear in various colors at different times,
such as orange or pink during sunrise and sunset, or gray when it's overcast'

User question: 'What's the color of grass?'
Bot response: 'The color of grass is typically green. This green color comes from chlorophyll,
 the pigment that plants use for photosynthesis. However, grass can also appear in different s
hades, such as yellow or brown, depending on factors like health, water availability,'

^C
Exiting...
```

Reflections

- Benefits:
  - OpenAI TTS enhances the **natural quality of responses**.
  - Threaded mode allows **real-time conversation flow**.

- Challenges:
  - **Slight delay** in OpenAI TTS processing compared to gTTS.
  - Thread synchronization requires careful **resource management**.

- Trade-offs:
  - Use threaded mode for **real-time interactions**.
  - Use linear mode for **simpler environments** or **external speaker setups.**

Thanks.

Zhiyu (Zoey) Zhang