AI Voice Assistant Bot With GTTS

Zhiyu (Zoey) Zhang

Project Design

- **API Integration:**
  - **GPT-4o-mini** for text generation.
  - **Whisper API** for speech-to-text transcription.
  - **Google gTTS** for text-to-speech.
- Supports **threaded** and **linear** modes.
- Designed for conversational AI with **wake words**, **stop words**, and **graceful shutdown**.

Key Features

- **AI Models:**
  - **GPT-4o-mini** balances cost and performance.
  - **Whisper API** for accurate transcription.
  - **gTTS** for speech output.
- **Modes:**
  - **Threaded mode** for simultaneous processing.
  - **Linear mode** for environments prone to feedback.

Key Features

- **Logging:**
  - **Verbose mode**: Detailed logs for debugging.
  - **Default mode**: Clean question-and-answer display.
- **Improved Functionality:**
  - Added **initialize_flags()** for reliable restarts.
  - Prevents **playback audio** from being re-processed.

Core Functionalities

- `record_audio()`: Captures and queues audio input.
- `transcribe_forever()`: Converts audio to text using Whisper API.
- `reply()`: Generates GPT-based responses and plays them using gTTS.
- `get_completion()`: Handles GPT-4o-mini interactions with the latest message structure.

Sample Use Cases

Threaded Mode with verbose:

- Detailed logs with timestamps.
- Continuous conversation flow without repeated wake words.

```
(base) zhangzhiyu@shengchanligongju w10 % python sts_gtts_threads.py --verbose
Listening...
[2024-11-24 14:47:45.728262] Transcription result: 'Hey computer, what's the color of sky?'
[2024-11-24 14:47:45.728412] Wake word detected. Starting conversation mode.
[2024-11-24 14:47:45.728495] Processing input: ' whats the color of sky'
[2024-11-24 14:47:46.649742] User question: ' whats the color of sky'
[2024-11-24 14:47:46.649775] Bot response: 'The color of the sky is typically blue during the d
ay due to the scattering of sunlight by the Earth's atmosphere. However, it can also appear in
shades of gray, orange, pink, or red during sunrise and sunset, and can be various colors durin
g'
[2024-11-24 14:47:49.273349] Audio file 'reply.mp3' generated successfully.
[2024-11-24 14:47:49.545111] AudioSegment loaded successfully.
[2024-11-24 14:48:07.924740] Playback completed.
[2024-11-24 14:48:07.925103] Temporary audio file 'reply.mp3' removed.
[2024-11-24 14:48:11.681782] Transcription result: 'What's the color of grass?'
[2024-11-24 14:48:11.681833] Processing input: 'Whats the color of grass'
[2024-11-24 14:48:12.464012] User question: 'Whats the color of grass'
[2024-11-24 14:48:12.464058] 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.'
[2024-11-24 14:48:13.259959] Audio file 'reply.mp3' generated successfully.
[2024-11-24 14:48:13.652122] AudioSegment loaded successfully.
```

Sample Use Cases

Threaded Mode with verbose:

- End conversations with stop word “stop”.
- Ignore input when not in conversation mode.
- Gracefully shut down the program.

```
[2024-11-24 14:48:36.601293] Transcription result: 'Thank you, stop.'
[2024-11-24 14:48:36.601330] Stop word detected. Ending conversation.
[2024-11-24 14:48:41.427607] Transcription result: 'How's the weather today?'
[2024-11-24 14:48:41.427648] Ignoring input (not in conversation mode).
[2024-11-24 14:48:49.776801] Transcription result: 'Hey computer, what's the newest iPhone?'
[2024-11-24 14:48:49.776896] Wake word detected. Starting conversation mode.
[2024-11-24 14:48:49.776926] Processing input: ' whats the newest iPhone'
[2024-11-24 14:48:51.041399] User question: ' whats the newest iPhone'
[2024-11-24 14:48:51.041454] Bot response: 'As of October 2023, the newest iPhones are the iPho
ne 15 series, which includes the iPhone 15, iPhone 15 Plus, iPhone 15 Pro, and iPhone 15 Pro Ma
x. They were announced'
[2024-11-24 14:48:52.317668] Audio file 'reply.mp3' generated successfully.
[2024-11-24 14:48:52.683755] AudioSegment loaded successfully.
[2024-11-24 14:49:10.147344] Playback completed.
[2024-11-24 14:49:10.147740] Temporary audio file 'reply.mp3' removed.
[2024-11-24 14:49:13.544759] Transcription result: 'Thank you, stop.'
[2024-11-24 14:49:13.544798] Stop word detected. Ending conversation.
^C
Shutting down...
All threads terminated. Goodbye!
```

Sample Use Cases

Threaded Mode by default:

```
(base) zhangzhiyu@shengchanligongju w10 % python sts_gtts_threads.py
Listening...
User question: ' whats the color of sky'
Bot response: 'The color of the sky is typically blue during the day due to the scattering of s
unlight by the Earth's atmosphere. However, it can also appear in various shades, such as gray
during overcast conditions, orange or pink during sunrise and sunset, and even'

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

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

Sample Use Cases

Linear Mode with verbose:

- Detailed logs with timestamps.
- Continuous conversation flow.
- End conversations with stop word “stop”.
- Ignore input when not in conversation mode

```
(base) zhangzhiyu@shengchanligongju w10 % python sts_gtts_linear.py --verbose
Listening...
[2024-11-24 16:44:33.694479] Waiting for input...
[2024-11-24 16:44:38.505653] Captured audio.
[2024-11-24 16:44:39.374597] Wake word detected. Starting conversation mode.
[2024-11-24 16:44:39.374661] User input: ', what's the color of sky?'
[2024-11-24 16:44:40.836489] Bot response: 'The color of the sky is typically blue during the d
ay due to the scattering of sunlight by the Earth's atmosphere. However, it can also appear in
various shades, such as gray during overcast conditions, orange or pink during sunrise and suns
et, and even'
[2024-11-24 16:44:42.479059] Audio file 'reply.mp3' generated successfully.
[2024-11-24 16:45:01.891031] Playback completed.
[2024-11-24 16:45:01.893211] Waiting for input...
[2024-11-24 16:45:06.017724] Captured audio.
[2024-11-24 16:45:07.224251] User input: 'What's the color of grass?'
[2024-11-24 16:45:07.939777] Bot response: 'The color of grass is typically green. This green c
olor comes from chlorophyll, the pigment that plants use in photosynthesis to absorb sunlight.
However, grass can also appear in different shades, such as yellow or brown, depending on facto
rs like drought,'
[2024-11-24 16:45:09.425045] Audio file 'reply.mp3' generated successfully.
[2024-11-24 16:45:28.785849] Playback completed.
[2024-11-24 16:45:28.786360] Waiting for input...
[2024-11-24 16:45:31.699433] Captured audio.
[2024-11-24 16:45:32.619155] Stop word detected. Ending conversation.
[2024-11-24 16:45:32.619204] Waiting for input...
[2024-11-24 16:45:36.737304] Captured audio.
[2024-11-24 16:45:37.430855] Ignoring input (not in conversation mode).
[2024-11-24 16:45:37.430899] Waiting for input...
[2024-11-24 16:45:45.839642] Captured audio.
[2024-11-24 16:45:46.544670] Wake word detected. Starting conversation mode.
[2024-11-24 16:45:46.544712] User input: ', what's that date today?'
[2024-11-24 16:45:46.954921] Bot response: 'Today's date is October 3, 2023.'
[2024-11-24 16:45:47.376280] Audio file 'reply.mp3' generated successfully.
[2024-11-24 16:45:51.543593] Playback completed.
[2024-11-24 16:45:51.543961] Waiting for input...
[2024-11-24 16:45:55.177780] Captured audio.
[2024-11-24 16:45:58.322571] Stop word detected. Ending conversation.
[2024-11-24 16:45:58.322634] Waiting for input...
^C
Exiting...
```

Sample Use Cases

Linear Mode by default:

```
(base) zhangzhiyu@shengchanligongju gtts % python sts_gtts_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 Earth's atmosphere. However, it can also appear in various shades, such as gra
y during overcast conditions, orange or pink during sunrise and sunset, and even'

User question: 'What's the color of the grass?'
Bot response: 'The color of grass is typically green, although it can vary in shade depending
on the type of grass, the season, and environmental conditions.'

^C
Exiting...
```

Reflections

- **Benefits:**
  - Threaded mode enhances **responsiveness**.
  - Linear mode ensures **clean audio** in challenging environments.
- **Challenges:**
  - Managing **audio feedback** in threaded mode with external speakers.
  - Complexity of **thread synchronization**.
- **Trade-offs:**
  - Choose threaded mode for **real-time interactions**.
  - Choose linear mode for **audio stability**.

Thanks.

Zhiyu (Zoey) Zhang