**Milestone-02 Documentation**

**Overview**

Milestone-02 builds upon the foundation set in Milestone-01 by introducing **Sentiment, Tone, and Intent Analysis** for audio inputs. This milestone integrates advanced AI capabilities to analyze both the transcribed content and vocal characteristics of user inputs, enabling the Real-Time AI Sales Assistant to provide more contextually accurate and empathetic responses.

**Modules and Functions Used**

**1. audio\_record**

* **Purpose**: Records user speech and converts it to text, while also saving the audio file for further analysis.
* **Source**: Imported from assignment\_02.

**2. text\_to\_speech**

* **Purpose**: Converts AI responses to speech for voice-based interaction.
* **Source**: Imported from assignment\_02.

**3. text\_response**

* **Purpose**: Generates AI responses based on user queries, incorporating sentiment and tone analysis.
* **Source**: Imported from assignment\_03.

**4. analyze\_audio**

* **Purpose**: Analyzes audio files for sentiment, tone, and intent using Google Generative AI.
* **Source**: Imported from Milestone\_02.

**New Features in Milestone-02**

1. **Audio Analysis**: Extracts and interprets user emotions, intentions, and vocal tones.
2. **Enhanced Context**: The AI Sales Assistant now leverages audio analysis insights to deliver more tailored and empathetic responses.
3. **Detailed Summary**: Provides structured feedback on the sentiment, tone, and intent of user speech.

**Functional Workflow**

**1. Enhanced Audio Input Processing**

* The audio\_record function records user input and returns:
  + Transcribed text.
  + A saved audio file for analysis.

**2. Audio Analysis**

* The analyze\_audio function processes the saved audio file using Google Generative AI:
  + Identifies **Sentiment** (Positive, Negative, Neutral).
  + Detects **Tone** (e.g., Excited, Frustrated, Curious).
  + Determines **Intent** (e.g., Seeking Information, Providing Feedback, Expressing Gratitude).

**3. Contextual Response Generation**

* The assistant combines:
  + Transcribed text.
  + Audio analysis summary (sentiment, tone, and intent).
* Passes this enriched context to text\_response for generating a refined reply.

**Functional Steps**

1. **Start the Program**:
   * Displays a welcome message to the user.
2. **Record and Analyze Audio**:
   * Records user speech via audio\_record.
   * Transcribes speech to text.
   * Saves the audio file for analysis.
   * Sends the audio file to analyze\_audio for tone, sentiment, and intent extraction.
3. **Generate AI Response**:
   * Combines transcribed text and analysis summary.
   * Passes them to text\_response for generating a customized AI reply.
4. **Provide Feedback**:
   * Outputs AI response as text and speech.
   * Displays detailed audio analysis results.
5. **Exit Condition**:
   * Ends the session if the user says "exit."

**Example Interaction**

**Input**

* User says:  
  *"I'm not happy with the delay in processing my order. Can you help me?"*

**Output**

* **Audio Analysis**:
* Sentiment: Negative
* Tone: Frustrated, Concerned
* Intent: Asking for help
* **AI Response**:
* "I understand your concern and apologize for the delay. Let me check your order status and provide you with a resolution immediately."

**Advantages of Milestone-02**

1. **Emotionally Intelligent Responses**: Enhances user experience by adapting to emotional cues.
2. **Rich Contextual Understanding**: Uses sentiment, tone, and intent to improve response relevance.
3. **Professional Interaction**: Delivers empathetic and context-aware replies.

**Limitations**

1. **API Dependency**: Relies on Google Generative AI for analysis.
2. **Processing Time**: Audio analysis may slightly delay response generation.
3. **Speech Quality Sensitivity**: Noisy or unclear audio may affect analysis accuracy.

**Conclusion**

Milestone-02 introduces advanced audio analysis capabilities, significantly enhancing the Real-Time AI Sales Assistant's ability to understand and respond to user emotions and intentions. This milestone lays the groundwork for building a deeply engaging and professional user interaction system. Future improvements may focus on optimizing processing speed and supporting multi-language audio inputs.