

User Guide: Medical Speech Translator

1. Introduction

The Medical Speech Translator is a web application, powered by advanced artificial intelligence and real-time speech processing, it transcribes spoken medical conversations, enhances terminology with precision, and translates them into your chosen language—all in real time.

2. Getting Started

2.1 System Requirements

- **Browser:** A modern web browser such as Google Chrome or Mozilla Firefox, with support for the Web Speech API.
- **Hardware:** A device with a microphone for audio input.
- **Internet:** Stable connectivity for AI-powered enhancement and translation via the Gemini API.
- **Permissions:** Grant microphone access when prompted by your browser.

2.2 Accessing the Application

1. Navigate to the application URL provided by your administrator.
2. Allow microphone access when prompted to enable speech recognition.

3. Interface Overview

The Medical Speech Translator features a user-centric interface designed with Material-UI and glassmorphism aesthetics. Key sections include:



- **Header:** Displays the application title ("Medical Speech Translator") and a theme toggle for switching between light and dark modes.
- **Language Selectors:** Dual dropdowns for selecting input (spoken) and output (translated) languages, accompanied by a swap button for seamless language reversal.

- **Controls Panel:** Offers intuitive buttons to manage speech recognition (start, stop, pause), play translated text, and reset transcripts, with real-time audio level feedback.
- **Transcript Panel:** A tabbed interface showcasing the original (enhanced) transcript and the AI-translated text, with interactive options for copying, speaking, and resetting.
- **Notifications:** Snackbar alerts provide feedback on actions or errors.
- **Loading Overlay:** A subtle backdrop with a progress indicator appears during initialization or AI processing.

4. How to Use

Follow these steps to harness the full potential of the Medical Speech Translator:

4.1 Configure Languages

- **Input Language:** Select the language you will speak from the "Input Language" dropdown (e.g., English - ). This determines the speech recognition language.
- **Output Language:** Choose the desired translation language from the "Output Language" dropdown (e.g., Spanish - .
- **Swap Languages:** Click the swap button (\leftrightarrow) to instantly reverse input and output languages, ideal for dynamic conversations.
- **Supported Languages:** Choose from 12 languages, including English, Spanish, French, German, Chinese, Japanese, Arabic, and more, each accompanied by a flag for visual clarity.

4.2 Record and Transcribe Speech

1. Click the **Start** button (featuring a microphone icon) in the Controls panel to initiate real-time speech recognition.
2. Speak clearly and at a steady pace, focusing on medical terminology if applicable.
3. Observe the audio level bar in the Controls panel to ensure optimal input volume.
4. The application will:
 - Transcribe your speech in real-time, displaying interim results.
 - Enhance medical terms (e.g., correcting "penacillin" to "Penicillin") using Gemini AI.
 - Translate the enhanced transcript into the output language.

4.3 View and Interact with Transcripts

- **Original Transcript Tab:** Displays the enhanced transcript in your spoken language, with corrected medical terms for accuracy.
- **Translated Transcript Tab:** Shows the AI-translated text in the output language, preserving context and medical precision.
- **Interactive Options** (available in both tabs):
 - **Copy:** Click the copy icon to save the transcript to your clipboard for documentation or sharing.
 - **Speak:** Press the speaker icon to hear the transcript read aloud in the respective language, using browser-based text-to-speech.
 - **Reset:** Select the replay icon to clear both transcripts and start a new session.
- **Navigation:** Switch between tabs to compare original and translated text, with smooth transitions powered by Material-UI's Fade component.

4.4 Manage Speech Recognition

- **Pause:** Click the pause button to temporarily halt transcription without ending the session, useful for brief interruptions.
- **Stop:** Press the stop button to finalize the transcription session, preserving the current transcripts.
- **Status Feedback:** Monitor the status chip in the Controls panel for updates like "Listening...", "Enhancing...", "Translating...", or "Idle".

4.5 Theme Customization

- Toggle between light and dark modes using the sun/moon icon in the Header. The dark mode (with a soothing #6fbd7f primary color) is ideal for low-light environments, while the light mode (#072c52) offers crisp visibility.

5. Advanced Tips for Optimal Use

- **Optimize Audio Input:**
 - Use a high-quality microphone in a quiet environment to minimize background noise.
 - Monitor the audio level bar (0–100 scale) to ensure your voice is detected clearly.

- **Handle Long Conversations:**
 - Pause recognition periodically to allow AI processing to catch up, especially for complex medical discussions.
 - Reset transcripts before starting a new topic to maintain clarity.
- **Language Selection:**
 - Choose specific dialects (e.g., en-US vs. en-UK) for better speech recognition accuracy.
 - Swap languages frequently in multilingual settings to adapt to different speakers.
- **AI Processing:**
 - Allow 1–3 seconds for Gemini AI to enhance and translate text, especially for longer phrases.
 - If translation fails (e.g., due to API issues), the original transcript remains available as a fallback.
- **Accessibility:**
 - Use the copy feature to share transcripts with colleagues or patients via text or email.
 - Leverage text-to-speech for auditory confirmation of transcript accuracy.

6. Troubleshooting

- **Microphone Issues:**
 - **Problem:** “Microphone permission denied” error.
 - **Solution:** Check browser settings to allow microphone access and reload the page.
- **Speech Recognition Failures:**
 - **Problem:** “No speech detected” or “Audio capture error”.
 - **Solution:** Ensure your microphone is functional and speak closer to it. Test in a quieter environment.
- **Translation Delays or Errors:**
 - **Problem:** Translated transcript shows “[Translation Failed]” or delays.
 - **Solution:** Wait a few seconds for AI processing. If persistent, check internet connectivity or contact the administrator to verify the Gemini API key.
- **UI Responsiveness:**
 - **Problem:** Buttons are disabled or unresponsive.
 - **Solution:** Wait for ongoing processes (e.g., enhancement, translation) to complete, indicated by “Enhancing...” or “Translating...” statuses.