Real-Time Assistive Speech Captioning System Architecture

**Client Side
(Web Browser)**

*Stage 1:
Capture*

**Server Side
(Flask + Whisper)**

*Stage 2:
Transcription*

Audio Capture
Web Audio API

Audio Reception
Base64 Decode

PCM Processing
Chunk Buffering

Format Conversion
(WebM→WAV)

Stutter Cleaning
Rule-based

*Stage 3:
Cleaning*

Base64 Encoding

Whisper ASR
Transcription

Next Sentence
Prediction

Audio Data

WebSocket
Transmission

Stutter Detection
Pattern Matching

*Stage 4:
Display*

**Output Display
(Web Interface)**

Raw Transcription

Cleaned Text

Stutter Patterns

Predictions