Memotag – Cognitive Decline Analysis System Report

Submitted by Kripa Mary Jose
2348526

This system is an intelligent application that seeks to assess potential cognitive decline by examining speech. This system integrates realtime transcription, cognitive metric analysis, sentiment analysis, and visualizations into a responsive web application. It offers a comprehensive solution for early detection and monitoring of cognitive health indicators through audio input.

Core Features

Audio Input Handling

- Supports both audio file uploads and live microphone recordings
- Automatic audio format conversion using FFmpeg
- Compatible with multiple formats including WAV and MP3

Speech Processing

- Speech To Text conversion using Google Speech Recognition API
- Accurate transcription display
- Error handling for unclear or low quality speech input

Cognitive Analysis

- Calculation of speech rate (words per minute)
- Detection and analysis of pauses and silence patterns
- Voice quality and pitch analysis (mean and variability)
- Generation of a cognitive risk score based on extracted speech features

Sentiment Analysis

- Sentence level sentiment classification
- Confidence scores for each sentiment prediction
- Aggregated summary statistics of positive, neutral, and negative sentiments
- Visualization of sentiment trends over time

Visualizations

- Display of audio waveform
- Generation of mel spectrograms
- Visualization of pitch contours
- Sentiment analysis trend graphs
- Color Coded cognitive risk indicators

Technical Implementation

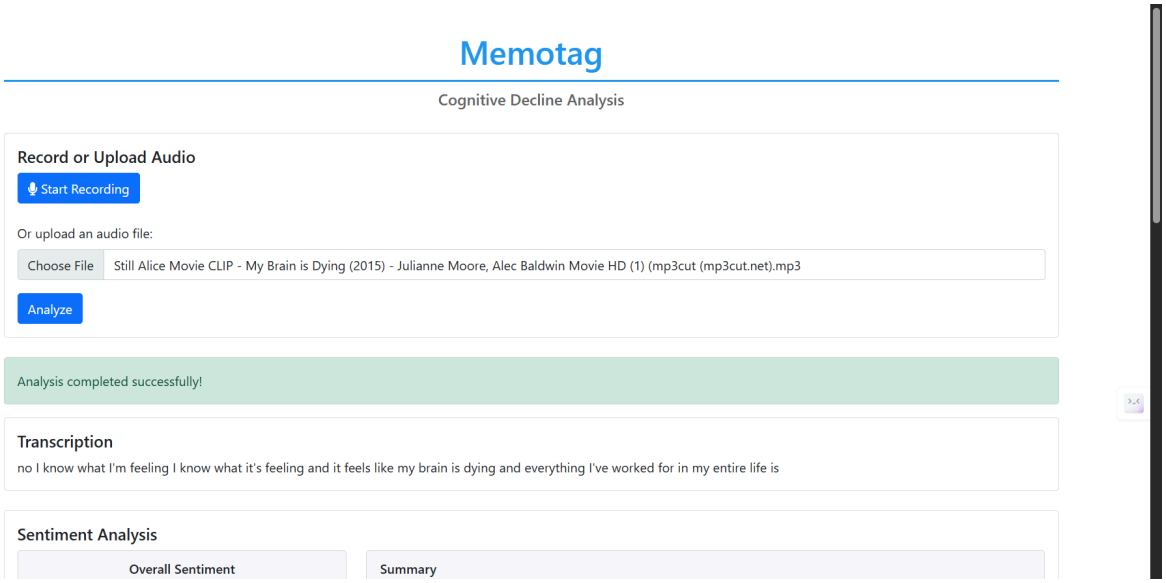
- Built with Flask as a RESTful API
- FFmpeg integration for audio preprocessing
- Feature extraction using Librosa
- Sentiment analysis using a pre trained BERT model
- Text processing via NLTK
- Visualization generation using Matplotlib
- Structured error handling and logging mechanisms
- Developed using HTML and Bootstrap for a modern, responsive design
- Supports real time recording and audio file input
- Displays transcription, analysis results, and visualizations dynamically
- Clean layout and intuitive navigation for a professional user experience

Analysis Workflow

In the Input Phase, users can either upload an audio file or record speech directly through the browser's microphone. Once received, the audio is normalized and converted into a compatible format using FFmpeg to ensure consistent processing. During the Processing Phase, the system transcribes the speech using Google Speech Recognition, extracts cognitive features such as speech rate, pauses, and pitch variation, and performs sentiment analysis on the transcribed text. Simultaneously, it generates relevant visualizations to support the analysis. In the Output Phase, the application displays the transcription results, provides a comprehensive summary of both sentiment and cognitive analysis, and presents visual summaries including waveforms, spectrograms, sentiment trends, and cognitive risk assessments for user interpretation.

Detailed Feature Breakdown

- Speech Analysis
- Words per minute and speaking rate computation
- Pause and silence frequency detection
- Pitch variability and tone analysis
- Assessment of voice modulation and clarity
- Calculation of cognitive decline risk score
- Classification into risk levels: Low, Medium, High
- Comprehensive interpretation of score with supporting metrics
- Interactive and interpretable plots such as Waveform,Spectrogram,Pitch Contour and Sentiment Trends
- Visual differentiation of normal and risky patterns
- Format validation and conversion fallback mechanisms
- Quality checks and user alerts for incompatible files
- Fallback methods for transcription failure
- Sentiment analysis exception handling
- Visualization error recovery and user notifications



Sentiment Analysis

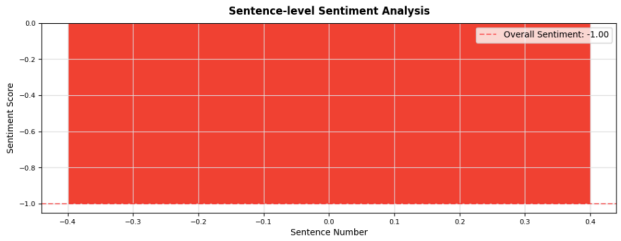
Overall Sentiment

-100%

Negative

Summary

0	0	1
Positive	Neutral	Negative



Sentence-level Analysis

no I know what I'm feeling I know what it's feeling and it feels like my brain is dying and everything I've worked for in my entire life is

Sentiment: -100.0% (Confidence: 23.7%)

Risk Assessment

Risk Level: Low

Sentiment: -100.0% (Confidence: 23.7%)

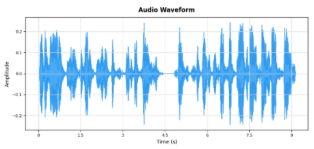
Risk Assessment

Risk Level: Low

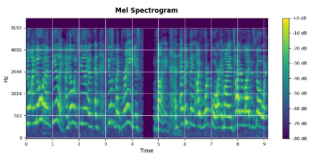
Normal cognitive function

Speech Analysis Visualizations

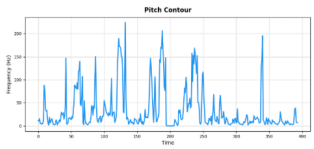
Waveform



Mel Spectrogram



Pitch Contour



Cognitive Analysis Metrics

Speech Rate

190.56 words/minute

Abnormal

Pause Analysis

Pause Ratio: 22.89%

Normal

Pitch Metrics

Mean: 1695.29 Hz

Normal

Voice Quality

ZCR: 0.1183

Abnormal