Speech Doctor

MVP Requirements (business requirement)

1. - "Submit a audio - in command line - output the number of pauses, Pause start(s) timestamps and respective duration
2. Sentence start, sentence end - time stamps
3. Filler words/Stuttering time stamp - what the filler word was - duration of words.
   Pick relevant model available with lightweight licensing such as MIT to select one which detects { oh, um, uh, er, ah} & { like, well, so, right, literally, okay } with better than 80% accuracy. Get on to data from public platforms (youtube etc) and data from the https://ai4bharat.org/models
4. Use a large Language Model - Preferably one forhttps://ai4bharat.org/models
   a. If model can determine the tone of sentence
   b. some type of more information on the quality of diction - diction profile etc
   c. Any other relevant information that can be told about the quality of diction - word speed, sentence speed, Tone of voice etc etc

Technically need following - Input is a Audio file - 30 sec to 200 sec duration

1. Start Session
2. Start processing audio
3. Apply noise reduction
4. Apply VAD classify word or no-word
   Objectives : Evaluate VAD - both classical DSP VAD and ML based VADs to detect pauses in audio clips with time stamps , duration , sentence start, sentence end
   Gather datasets for VAD evaluation
   Technical Objective - To define and create a C API / python which can be invoked from either C/C++ , Python or C# wrappers. Results will be returned in JSOn format. Please note speed is of essence. All processing should be done in a few seconds.
5. Use a pre-trained ML model, which are proven in south asian accents, to detect legible language(only English to start with) words. Refine classification
6. Store the time stamps of the events
7. Update the session information to dB
8. Close session

To evaluate different models available with lightweight licensing such as MIT to select one which detects { oh, um, uh, er, ah} & { like, well, so, right, literally, okay } with better than 80% accuracy.
Gather datasets for Filler words/Stuttering evaluation

Integrate VAD
Integrate the Filler words/Stuttering model