**Automatic Answer Grading**

**Video to Audio using augmentation technique**:

Processing audio extracted from a video file. It uses libraries like moviepy and soundfile to extract audio from a video, measure its loudness, and normalize it to -12 dB LUFS(Loudness Units Full Scale). The code also demonstrates saving the modified audio as a new WAV file.

**Selecting the best model for speech-to-text transcription.**

**{Audio to text using Whisper Models:**

Small, medium, large, large-v2

calculate Word Error Rate (WER), BLEU score, and Levenshtein distance between two strings, using the NLTK library.

(Accuracy of machine translation or text generation models. For WER and Levenshtein distance: Lower values indicate better accuracy.

For BLEU score: Higher values indicate better accuracy.)}

**WhisperX\_Speaker\_Diarization using large-v2 Model:**

WhisperX library transcribed an audio file, segmenting the text by speakers, and then grouping continuous segments spoken by each speaker.

[openai/whisper-large-v2 · Hugging Face](https://huggingface.co/openai/whisper-large-v2)

**why we used Whisperx, why not whisper or something**

**WhisperX,** an extension of the Whisper model, offers several features for time-accurate speech transcription, including**:**

**1. Speaker Recognition**: It recognizes different speakers and makes them identifiable in the transcribed speech text.

**2. Word-Level Timestamp Accuracy:** WhisperX provides accurate word-level timestamps for utterance-level detection, improving the quality of matching and preventing timestamp errors**.**

**3. Diarization:** The system includes speaker diarization, which is the process of partitioning an audio stream into homogeneous segments according to the speaker identity**.**

**4. Batch Inference and Transcription:** WhisperX enables batched whisper transcription, resulting in improved accuracy and efficiency. It also offers batch inference within a file**.**

**5. Reduced Hallucinations:** It reduces hallucinations in speech detection, leading to higher transcription accuracy**.**

**6. No-Speech Detection:** WhisperX includes a feature for no-speech detection, which can be beneficial for various applications**.**

These features make WhisperX suitable for various applications, including speech-to-text conversion for different use cases.

Links used: [m-bain/whisperX: WhisperX: Automatic Speech Recognition with Word-level Timestamps (& Diarization) (github.com)](https://github.com/m-bain/whisperX)

**Evaluation:**

Evaluated using OpenAI, gpt-3.5-turbo model

**For Grading Text-Based Responses:**

1. Fluency
2. Clarity
3. Confidence
4. Enthusiasm
5. Unique qualities
6. Professionalism
7. Engagement
8. Structured Answers
9. No Fillers
10. Focused
11. Authentic
12. Overall

**Accuracy and performance of your evaluation system**

**Mean Absolute Error (MAE)** (Lower values indicate better performance)

**Root Mean Squared Error (RMSE)** (Lower values indicate better performance)

**Pearson Correlation Coefficient** (a higher Pearson correlation coefficient indicates a stronger linear relationship between human and AI scores, suggesting that the AI grading system is more closely aligned with human grading for that particular parameter.)

**Correlation with Human Judgments**

This metric measures the correlation between AI-generated marks and human judgments, providing insights into the agreement between the two sets of marks

**Visualization graphs**: Histogram and violin graph showing the frequency of marks of different parameters, comparing with

Correlation bar graphs for every column and average of all columns

**Automatic Answer Grading with Speech Recognition and Evaluation**A system for automatically grading spoken answers using speech-to-text conversion and evaluation metrics.**Stages:**

1. **Preprocessing**:Extract audio from video using libraries like moviepy and soundfile.Normalize audio loudness to -12 dB LUFS.
2. **Speech-to-Text Transcription**:Choose the best model (e.g., Whisper small, medium, large, large-v2) based on metrics like WER (Word Error Rate), BLEU score, and Levenshtein distance.Transcribe the audio using the chosen model with Whisper library.
3. **Speaker Diarization:**Use WhisperX\_Speaker\_Diarization with the WhisperX library for speaker identification and segmentation. This offers benefits like:Speaker recognitionWord-level timestampsSpeaker diarizationBatch inferenceReduced hallucinationsNo-speech detection
4. Evaluation:Use OpenAI's gpt-3.5-turbo model to evaluate the transcribed text based on predefined criteria:FluencyClarityConfidenceEnthusiasmUnique qualitiesProfessionalismEngagementStructured AnswersNo FillersFocusedAuthenticOverall
5. Performance Measurement:Calculate Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to assess the deviation between AI and human graders.Determine the Pearson Correlation Coefficient to measure the linear relationship between human and AI scores for each parameter. This indicates agreement between the two.
6. Visualization:Generate histograms and violin plots to show the frequency distribution of marks for each parameter from both AI and human graders.Create correlation bar graphs to visualize the correlation between human and AI scores for each parameter and an overall average.
7. Benefits:Saves time compared to manual grading.Offers speaker identification for group responses.Provides detailed evaluation metrics for student feedback.
8. Considerations:Accuracy of speech recognition models.Potential bias in the evaluation criteria.Applicability to different answer types (factual vs. creative).This system provides a framework for automating answer grading with speech recognition and evaluation. Further research could focus on improving accuracy and incorporating human feedback for better calibration.