**Interview Performance Evaluation Report: Audio**

**Objective:**The goal is to evaluate interview performance based on audio features such as modulation, confidence, and filler word usage. This evaluation provides insights into the quality of speech, confidence levels, and language efficiency.**Methodology:Audio Processing and Feature Extraction**:Modulation Analysis: Measures variations in pitch, volume, and tone within the speech. High modulation scores indicate greater fluctuations in these features, reflecting dynamic and engaging speech, whereas low scores suggest monotony or lack of expressiveness.

Confidence Assessment: Evaluates speech rate, pause frequency, and pitch stability. High confidence scores are indicative of fluent, steady speech with minimal pauses, suggesting greater speaker confidence. Conversely, low scores suggest frequent pauses and unstable pitch, indicating lower confidence.

Filler Words Detection: Identifies the frequency of filler words such as "um," "uh," and "like" within the speech. A higher ratio of filler words indicates less clarity and organization in speech, which can negatively impact communication effectiveness.

**Emotion Analysis**:**Emotion Classification**: Uses deep learning models to classify emotions from facial images or audio signals. Each emotion, such as fear, anger, disgust, neutrality, sadness, happiness, or any other relevant emotion, is assigned a weight based on its significance.

1. Fear: Could indicate nervousness, which might negatively impact the score
2. Happy: May suggest confidence and a positive attitude, which could positively impact the score
3. Disgust: Might be a sign of discomfort or disagreement, potentially impacting the score negatively
4. Neutral: Indicates a balanced and calm response, which might be neutral or slightly positive.
5. Angry: Could reflect frustration or hostility, negatively impacting the score.
6. Sad: May indicate a lack of confidence or distress, negatively impacting the score
7. Ps(Pleasant Surprise): Could indicate engagement orinterest, positively impacting the score

**Grade Calculation**: The total score is derived from the sum of weights assigned to detected emotions. This score is then normalized to a scale of 0 to 100, reflecting the emotional intensity and expressiveness of the speech.

**Normalization of Scores**:Modulation Score: Normalized on a scale of 0 to 1 by dividing the raw score by 350. This ensures scores are within a consistent range for comparative purposes.Confidence Score: Normalized on a scale of 0 to 1. This normalization ensures that the confidence score aligns with the expected range.Filler Words Ratio: Inverted so that a lower ratio (fewer filler words) results in a higher normalized score, aligning with the goal of reducing filler usage for clearer speech.Emotion Grade: The grade score derived from emotion analysis is also normalized to fit into the overall evaluation framework.

**Combining Scores**:The final interview score is a weighted combination of the normalized modulation, confidence, and filler word scores, along with a grading score based on predicted labels. This composite score reflects overall performance, with weights assigned to each component to reflect its importance in the evaluation.

Technologies Used:**Librosa**: A Python library for audio and music analysis, used for feature extraction such as pitch tracking, volume analysis, and onset detection.**Speech Recognition**: A tool for transcribing spoken words into text, enabling the detection of filler words and assessment of speech clarity.

Emotion Detection Model: **LSTM** for classifying emotions from audio signals

Conclusion:This approach provides a comprehensive evaluation of interview performance by analyzing various aspects of speech and emotional expression. The normalization and combination of scores ensure that different features are assessed uniformly and contribute to a holistic performance score. This evaluation method helps in understanding the effectiveness of speech delivery, confidence level, and language efficiency and emotional engagement in interviews.

**Speech Emotion Recognition**

**Dataset Information**

There are a set of 200 target words spoken in the carrier phrase "Say the word \_" by two actresses (aged 26 and 64 years). Recordings were made portraying each of seven emotions: anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral. There are 5600 data points (audio files) in total.

The dataset is organized such that each of the two female actors and their emotions are contained within its own folder. Within that, all 200 target words' audio files can be found. The format of the audio file is WAV.

**Output Attributes**

* **anger**
* **disgust**
* **fear**
* **happiness**
* **pleasant surprise**
* **sadness**
* **neutral**

**Download Links**

* Dataset: [Toronto Emotional Speech Set (TESS)](https://www.kaggle.com/ejlok1/toronto-emotional-speech-set-tess)

**Libraries**

Pandas, matplotlib, keras, tensorflow, librosa

**Neural Network**

* LSTM Network

**Achieved Accuracy**

**100.00%**

Things to do:

**Evaluation:**

Ground Truth Comparison

Performance Metrics:

Correlation Analysis