**PROPOSED TECHNIQUE FOR LGSI PROJECT**

**AI-based audio/video quality checking - solution to measure the quality of audio/video**

Animesh Raj Ponnuri Aniruddha Zayed Haque

RA2112704010005 RA2112704010015 RA2112704010010

CSE DATA SCIENCE CSE DATA SCIENCE CSE DATA SCIENCE

**Objective For the Project**

The intention is to create a solution that can automatically analyze audio and video files and provide a quality assessment.

**Proposed Solution**

The proposed solution leverages the applications of machine learning and deep learning to analyze audio and video files automatically and provide accurate quality assessments. The system can provide robust evaluations by utilizing innovative techniques for feature extraction and multi-modal fusion, ensuring improved user experience and satisfaction with multimedia content.

**Audio Quality Checking:**

A novel model for audio quality assessment is presented, incorporating a powerful combination of Bidirectional Long Short-Term Memory (BiLSTM) and an Attention Mechanism. The BiLSTM emulates human auditory perception, effectively capturing comprehensive information from the audio recording. Simultaneously, the Attention Mechanism enhances discrimination by highlighting relevant target-related features, effectively separating desired signals from potential interferences. This joint utilization of BiLSTM and Attention Mechanism creates a sophisticated framework that significantly improves the accuracy and robustness of audio quality assessment, offering valuable insights into the perception and discrimination of audio content, thus contributing to a more refined and reliable evaluation of audio quality.

**Video Quality Checking:**

A new approach, the Divide and Conquer Video Quality Estimator (DCVQE), is proposed for No-Reference Video Quality Assessment (NR-VQA). The novel Divide and Conquer Transformer (DCTr) architecture initiates by extracting frame-level quality embeddings from the input video. Two types of Transformers are then introduced to gradually learn the clip-level and video-level quality embeddings within the DCTr layer. Through the stacking of multiple DCTr layers and the addition of a regressor, a hierarchical model called the DCVQE is constructed, enabling accurate prediction of the input video's quality score. Furthermore, an innovative correlation loss term is suggested, considering the order relationship among the training data, to effectively guide model training.

**Requirements**

The proposed research on audio and video quality assessment requires access to large, high-quality datasets for effective model training. The TIMIT dataset and VoxCeleb dataset are identified as valuable resources for the audio quality assessment project, offering diverse audio recordings from different speakers to train the BiLSTM and Attention Mechanism model. For the video quality assessment project, datasets like LIVE Video Quality Database and Ultra Video Group's Video Quality Database are deemed suitable, containing video clips with varying quality levels to train the Divide and Conquer Video Quality Estimator (DCVQE). Python will be the primary language, supplemented by deep learning libraries like TensorFlow, PyTorch, or Keras for model implementation. Additionally, NumPy, Pandas, Scikit-Learn, Librosa, OpenCV, and FFmpeg will be utilized for data and audio/video processing, enabling efficient data preparation and result interpretation.

**Conclusion**

In conclusion, The proposed AI-based audio and video quality-checking solutions incorporate novel models - BiLSTM with Attention Mechanism for audio and DCTr for video. These advanced frameworks enhance accuracy and offer valuable insights, ensuring refined evaluations.