International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

A Review on Existing Systems in Audio & Interview Analysis

Vaibhavi Sanap
Computer Engineering Department
Sinhgad College of Engineering (SPPU)
Pune,India
vaibhavi.r.sanap@gmail.com

Kiran Lavhale
Computer Engineering Department
Sinhgad College of Engineering (SPPU)
Pune,India
lavhalekiran46@gmail.com

Aditya Manchare
Computer Engineering Department
Sinhgad College of Engineering (SPPU)
Pune,India
adimachare17@gmail.com

Hardik Maind
Computer Engineering Department
Sinhgad College of Engineering (SPPU)
Pune,India
maindhardik@gmail.com

Mayuri Khade
Computer Engineering Department
Sinhgad College of Engineering (SPPU)
Pune, India
makhade.scoe@sinhgad.edu

***Abstract*—In the dynamic landscape of recruitment, the need for sophisticated yet equitable interview assessment tools is undeniable. The unpredictability inherent in interviews often leads to apprehension among candidates, potentially impeding their performance. In response, our project introduces a novel Interview Assessment Tool, integrating advanced machine learning techniques with intuitive interfaces to streamline interviews, gauge user confidence, and monitor progress. This tool offers a tailored experience, ensuring data security while upholding the integrity of assessments. By generating job-specific questions, it enhances the relevance of evaluations. The AI-Driven Assessment System provides real-time feedback through natural language processing and audio analysis, facilitating continuous improvement. The Feedback Mechanism delivers comprehensive reports, identifying strengths and areas for growth, fostering self-reflection and skill enhancement. The User Progress Tracking System enables users to monitor their development across multiple interviews. A notable innovation is the incorporation of voice analysis to assess user confidence, providing valuable insights for candidates and recruiters alike. Our Interview Assessment Tool revolutionizes interviews through AI and data-driven insights, delivering a secure, personalized, and equitable assessment experience, empowering candidates to bolster their confidence and excel in the competitive job market.**

Index Terms— Confidence, Audio-Based Evaluation, Mock Interviews, Speech Emotion Recognition, LSTM, MFCCs

INTRODUCTION

In the past, interviews were all about meeting face-to-face, but nowadays, they're often done online for convenience. This change has made things easier for both recruiters and candidates. Mock Interview Platforms have popped up to help candidates prepare, taking into account this new online trend. Even though there are plenty of esources out there to help candidates get ready, there's still work to be done in figuring out the best ways to assess and give feedback. These platforms are super useful because they let candidates see how well-prepared they are in a realistic setting. They look at things like behavior and non-verbal cues to see how ready candidates are for specific projects. This literature review dives into how current systems analyze these factors and tries to set the stage for a new "Audio-Powered Candidate Evaluation System Contextual & Grammar Analysis: In this evaluation, we employ a text-to-speech method. The audio input is transformed into text to scrutinize the candidate's choice of words. Subsequently, Natural Language Processing Techniques are applied to assess the response's accuracy and its sentence structure.

"

I. LITERATURE SURVEY

Literature Survey is summary of existing research work conducted on a particular topic. The present literature focuses on analyzing non-verbal cues such as head Motion [1], speech recognition [1],[2] and face recognition [2], to understand the body language of the user and to provide a feedback based on it.

The research talks about using different computer methods to understand speech better. They mainly focus on one thing: how high or low someone's voice sounds, which is called pitch.Voice recognition systems today rely on mostly classifying confidence based upon the quality of audio received[2].Our reference[2] focuses on emotion classification using the speech, classifying it according to various emotions such as sad, happy, nervous among others[2]. The study focuses on understanding emotions conveyed through speech. They use a technique called Mel-frequency Cepstral Coefficient to clean up the data and make it clearer. Then, apply a LSTM (Long Short-Term Memory) model to analyze the speech and classify emotions such as sadness, happiness, or nervousness.

Additional studies employ MLP classifiers to gauge overall confidence levels. These classifiers blend observations of non-verbal cues alongside speech recognition techniques to yield comprehensive results.
The Deep Learning based Self- assessment tool[3], takes in case a real world scenario for actual interview, while using Svm and CNN algorithms. One particular research undertook by Yi-Chi Chou, Felicia R. Wongso and others [4], holds the base for our project, analyses the input audio deeply, providing a comprehensive feedback based upon 3 parameters – Speaking rate, Frequency, and Amplitude of the voice. It also classifies personality types in a DISC frame (dominance, Influence, Steadiness, Compliance). This particular research is comprehensive to classify the personality types based upon inputs of audio-visual format.

This specific study employs the ARD regression model to measure and categorize personality types. Additionally, we delve deeper into current audio analysis methods to develop a robust approach for interpreting and scrutinizing audio data.The Audio classification approach by Krishna Kumar and Kapil Chaturvedi [5] lays the base for us, to classify audio based upon the features extracted from the input, using the MFCC, Spectrogram and relevant Python libraries which prove essential in audio analysis. This method utilizes the neural network model for the classification of audio, making it a good framework for classification.
The study detailed in paper [6] emphasizes the application of LSTM and deep neural networks for sorting audio data. It specifically concentrates on distinguishing various acoustic environments based on the audio input.The investigation detailed in the research paper [7] guided us in understanding the essential steps for feature extraction and audio classification. It underscores the significance of factors such as MFCCs, Pitch Frequency, Zero-Crossing rates, bandwidth, and sub-band energy in this process.

In our analysis, it becomes evident that existing behavioral analysis models incorporate a blend of non-verbal cues and speech frequency. However, this amalgamation leaves a void for further exploration in the realm of audio-based behavioral classification. By focusing solely on speech frequency and non-verbal cues, these models overlook the nuanced aspects of audio signals that could offer valuable insights into candidate behavior. This underscores the necessity for research initiatives aimed at harnessing the potential of audio data in enhancing the accuracy and comprehensiveness of behavioral assessments.
Moreover, our parallel investigation lays the groundwork for advancing audio-based classification methodologies, offering invaluable insights into the requisite strategies for implementing such models within the Interview Assessment Scenario. By elucidating the fundamental principles and techniques essential for audio-based classification, this research provides a solid foundation upon which to build innovative approaches that integrate audio analysis into the assessment process. This not only enriches the scope of behavioral analysis but also opens avenues for more nuanced and insightful evaluation of candidate performance.
.

Paper No	Algorithm	Accuracy
1	MLP	59.4%
1	K neighbourhood method	50.7%
1	Decision tree	52.2%
1	Naive Bayes	59.4%
3	CNN	70%

Table 1. Comparison of Accuracies of various Models implemented in papers.

II. PROPOSED SYSTEM

Keeping current systems in view, we propose a similar architecture,(refer diagram 1) but based on only the audio as an input for our system. The behavioral analysis have expanded to the classification of Confidence, lifting up from the previous emotion classification.

We also aim to provide a comprehensive feedback, highlighting the confidence, Subject Understanding of the user, with their grammar analysis to give a proper view of the user's current capabilities. As mentioned, we also plan to develop a User Feedback Progress Tracking System to track the user's progress to give them a clear analysis of their growth or strengths, providing a sense of confidence for them to face real Interviews.

The proposed Modules in our System are:

1. Questioning Module
2. Analysis Module
3. Feedback Mechanism & User Progress Tracking System.

The functioning of these modules is as follows:

1. *Questioning Module*:

In a mock interview scenario, the questions asked pertain to the role offered or the general understanding of the subject in case. The current models in application are found to be insufficient for comprehensive questioning, as they tend to either answer a limited number of questions, having same questions each time. We aim to overcome this, by creating a module which would ask questions based upon the different subjects covered in that domain, in addition to asking different questions most of times, to gauge the contextual understanding of the user's knowledge.

2. Analysis Module:

This module, is a product of all the audio research we undertook. The module is divided into three parts:

- Confidence Analysis
- Contextual Analysis
- Grammar Analysis

We aim to use various machine learning models and NLP techniques to analyze the classify the Audio gathered as the input.

3. Feedback Mechanism & User Progress Tracking System.

The system aims to offer users a profound insight into their strengths and capabilities by providing detailed feedback. This not only ensures a clear understanding of their performance but also stores historical feedback, enabling users to track their progress and enhancing overall confidence. The feedback mechanism serves as a valuable tool in elevating user confidence through comprehensive and continuous analysis.

III.IMPLEMENTATION

The Questioning Module comprises a dataset primarily consisting of frequently asked interview questions. However, our aim is to ensure its comprehensiveness by incorporating questions from various subjects within the domain, along with a random question selection mechanism. These questions will be dynamically presented on the screen through a randomized selection process from the question dataset.

The Audio Analysis Module is designed to function in a step-wise manner:

a) Audio Extraction: Utilizing Python Libraries such as Librosa and Pydub, the audio data will be converted into a usable format, eliminating any inconsistencies or noise present in the input. Subsequently, the MFCC method will be employed to extract the requisite data from the audio.

b) Confidence Analysis: The extracted audio will undergo analysis using a combination of LSTM and MLP classifiers to provide confidence estimations. Previous research has demonstrated the highest levels of accuracy with these models. MLP, which relies on mathematical functions, serves as the foundation for the MLP Machine Learning Model. These functions include:

1. Forward Propagation:

$$o_j^l = \sigma(\sum_i w_{ji}^l \cdot o_i^{l-1} + b_j^l)$$

2. Backward Propagation

$$\frac{\partial L}{\partial w_{ji}^l} = (\hat{y} - y) \cdot \sigma'(\sum_i w_{ji}^l \cdot o_i^{l-1} + b_j^l) \cdot o_i^{l-1}$$

3. Loss Function

$$L(y, \hat{y}) = \frac{1}{2}(\hat{y} - y)^2$$

c)Contextual & Grammar Analysis: In this evaluation, we employ a text-to-speech method. The audio input is transformed into text to scrutinize the candidate's choice of words. Subsequently, Natural Language Processing Techniques are applied to assess the response's accuracy and its sentence structure.

The User Feedback is derived from the results of these models and presented in an easy-to-understand format.

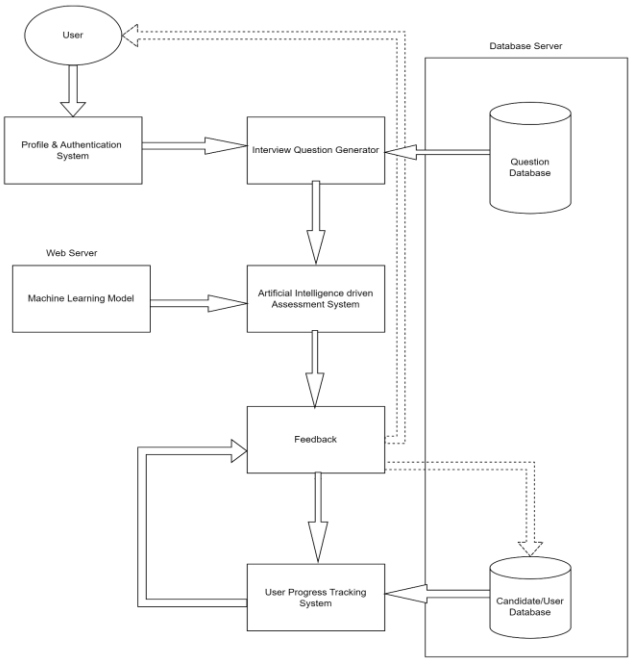

Diagram 1:System Overview Diagram

III. CONCLUSION

Accordingly, we have discerned a notable research void within the sphere of behavioral assessment and estimation, specifically honing in on the utilization of audio input. Our mission is to rectify this gap by crafting a system that not only addresses this deficiency but also grants access to historical data, thereby affording a comprehensive grasp of users' capabilities.

At its core, the system epitomizes a substantial leap forward in the arena of candidate evaluation. With its dynamic features seamlessly interwoven with robust feedback mechanisms, it endeavors to embolden candidates, fostering enhanced performance while alleviating the burden of stress and nervousness.

V. REFERENCES

[1] Erina Kasano, Shun Muramatsu, Akihiro Matsufuji, Eri Sato-Shimokawara , and Toru Yamaguchi- "Estimation of Speaker's Confidence in Conversation Using Speech Information and Head Motion"

[2] Rubi Mandal,Pranav Lohar,Dhiraj Patil, Apurva Patil, Suvarna Wagh, - "Interview-Based mock interview evaluator : An emotion and confidence classifier model"

[3Sumedh Anglekar,Urvee Chaudhari,Atul Chitanvis,Dr. Radha Shankarmani - "Deep Learning based Self-Assessment Tool for Personality Traits and Interview Preparations", IEEE 2021

[4]] Yi-Chi Chou, Feli cia R. Wongso, Chun-Yen Chao, Han-Yen Yu-"An AI Mock-interview Platform for Interview Performance Analysis"

[5] Krishna Kumar, Kapil Chaturvedi, -"An Audio Classification Approach using Feature extraction neural network classification Approach."

[6] Yanxiong Li, Xianku Li, Yuhan Zhang, Wucheng Wang, Mingle Liu, Xiaohui Feng-" Acoustic Scene Classification Using Deep Audio Feature and BLSTM Network"

[7] Bai Liang, Hu Yanli, Lao Songyang, Chen Jianyun, Wu Lingda – "Feature Analysis and Extraction for Audio Automatic Classification"