International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 15 Issue 3, March-2026

AI- Driven Mock Interview: A New Era In Candidate Preparation

Prof. Baliram Deshmukh[1], Shubham Yadav[2], Omkar Khaladkar[3], Karan Indalkar[4], Pranav Kare[5]

[1,2,3,4,5] Navsahyadri College Of Engineering, Pune, Maharashtra, India

ABSTRACT

The speed with which AI is developing has had a profound impact on people acquisition and recruiting. Using methods from audio analysis, machine learning, and natural language processing (NLP), this article details the development and launch of an AI-powered Interview System that can automatically assess job candidates. The proposed system conducts structured interviews, analyzes candidate responses in real time, and evaluates performance based on linguistic quality, semantic relevance, confidence indicators, and behavioral attributes.

The system integrates speech-to-text conversion, sentiment analysis, keyword extraction, and predictive scoring models to generate objective and data-driven assessments. By minimizing human bias and improving scalability, the AI-powered interview platform enhances efficiency, consistency, and fairness in candidate screening. Experimental evaluation demonstrates improved assessment accuracy and reduced recruitment time compared to traditional interview methods.

INTRODUCTION

The recruitment process is a critical function in every organization, directly influencing productivity, performance, and long-term success. Traditional interview methods rely heavily on human judgment, which is often influenced by unconscious bias, inconsistency, fatigue, and time constraints. As organizations scale and applicant volumes increase, manual interview processes become inefficient, costly, and difficult to standardize. These limitations create the need for intelligent, automated systems capable of conducting fair, consistent, and data-driven candidate evaluations.

A number of industries are seeing the revolutionary effects of artificial intelligence (AI), including medicine, banking, schools, and HRM. Automating tasks such as resume screening, applicant shortlisting, behavioral analysis, and predictive performance modeling is made possible in the recruiting process by AI. Artificial intelligence systems are able to sift through potential answers and derive deeper meaning, emotional tone, and communication quality by using methods like Natural Language Processing (NLP), Machine Learning (ML), and voice recognition.

The emphasis of this work was on evaluating interviews with Chinese participants since the research was lacking. A wide range of current approaches are available, including virtual reality interviews, AI chatbots, and in-person interviews. The use of the DISC personality model during interviews to explore applicants' character attributes was another area of emphasis. Researchers rely on the Big Five model of personality characteristics when conducting interviews with social media users in majority of their research [8].

The purpose of this research is to address the needs of both job hopefuls and recruiters in relation to video interviews. In light of this, we suggested a framework for automatic scoring of asynchronous video interviews. By using the video interview platform, both the employment agency and the candidate may save time, save costs, and overcome obstacles associated with remote work [9]. Candidates for jobs may use it to better understand their own habits, vocal inflection, and the topics covered in interviews. Recruiters might use the given result as a guide to choose applicants for the screening procedure.

That is the outline of the paper. Existing mock interview products and historical research on video interviews and personality analysis are briefly covered in Section II. In Section III, the methods and platform structure used in this research are shared. Conclusions are presented in Section IV. Thoughts and conversations about potential developments for the future are discussed in Section V.

RELATED WORK

A. Mock Interview

As a result of the epidemic, an increasing number of recruiters are including video interviews into their employment processes. In the United States, many people have utilized video interviewing sites like HireVue. From the answers given by candidates, the platform derives an assessment of their hard and soft talents, including their ability to work in a team and solve problems [10][11]. A candidate's "employability" score is also determined by how HireVue's AI analyzes their facial expressions, eye contact, and speech.

Aurora Cloud is a Taiwanese business that makes an AI interview system for people who speak Chinese. In order to decipher non-verbal cues such as micro expressions, speech rate, gestures, and audio elements, the platform employs audio-visual processing [12]. The applicants' Big Five personality characteristics were then predicted using these features [12]. One such product from Taiwan, Lasso AI Interview, uses a candidate's facial expressions, intonation, speaking speed, and mood fluctuations to determine their inherent features.

B. Interview Features

Observing and responding to nonverbal clues is just as crucial in interviews as speaking well. While words make up only 7% of communication, nonverbal cues like body language and vocal intonation contribute for 55% and 38%, respectively, according to the 7-38-55 rule. “Filming an interview in its entirety allows for the capturing of all of the interview's nonverbal cues, including posture, intonation, facial expressions, and more. The goal of these research is to identify these patterns of behavior so that interviewers may better gauge how well a candidate will do.

Researchers in one research suggested using structured video interviews with multimodal feature extraction to glean audio, visual, and textual information. Some examples of the multimodal traits include head position, eye movement, Action Units, various auditory characteristics, and textual clues. Predicting hiring recommendation ratings using both text and auditory clues was successful. Each interview's prosodic, lexical, and facial characteristics (such as the strength of the grin and the body language) were recorded in a separate research. Recommendation ratings and behavioral trait predictions were both based on these characteristics.

C. Personality Traits Analysis

The degree to which an individual enjoys and excels in their work is highly dependent on their personality features. Unfortunately, interviewers aren't always accurate when it comes to gauging a subject's character attributes in a little period of time. Researchers were able to deduce a candidate's communication abilities and Big Five personality attributes from their facial expressions. While the model performed well in predicting neuroticism, openness, agreeableness, and communication skills, it failed miserably when it came to extraversion and conscientiousness.

Predicting personality characteristics is also greatly influenced by textual clues. Using the median values of the Big Five personality characteristics as a criterion, a study suggested classifying each attribute as high or low. Next, the textual characteristics extracted from the video interviews were categorized according to each personality attribute using a binary classification model.

OUR PROPOSED METHOD

The researchers in this study came up with the idea of an automated grading system for asynchronous video interviews. In order to score candidates, we use characteristics extracted from their video interviews to assess their personality traits and how well they did throughout the interview. Using visual aids like charts and graphs, the software generates a quantitative report in real time.

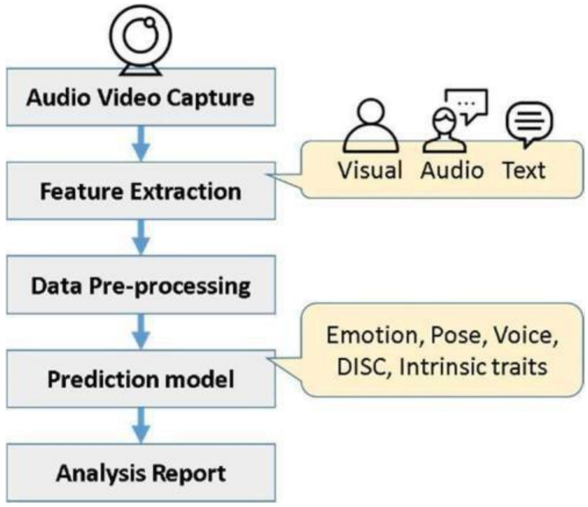


Fig. 1. The MIP framework

The MIP's structure is shown in Fig. 1. Initially, participants will be asked to film and upload a video interview. The platform's online interface allows users to practice interviewing with human resources or supervisors from the organization. After the user submits a video, the platform will preprocess the data it has retrieved by looking for visual, auditory, and textual elements. The next step is to run the data through five different prediction models: DISC, intrinsic qualities, emotion, voice, and posture. Finally, a quantitative report showcasing the outcomes of the study and predictions is generated by the platform.

A. Data Description

1. Data Collection

One hundred native Chinese speakers from a variety of professions were recruited to take part in the study. Subjects were asked to sit in silence for the practice interviews.

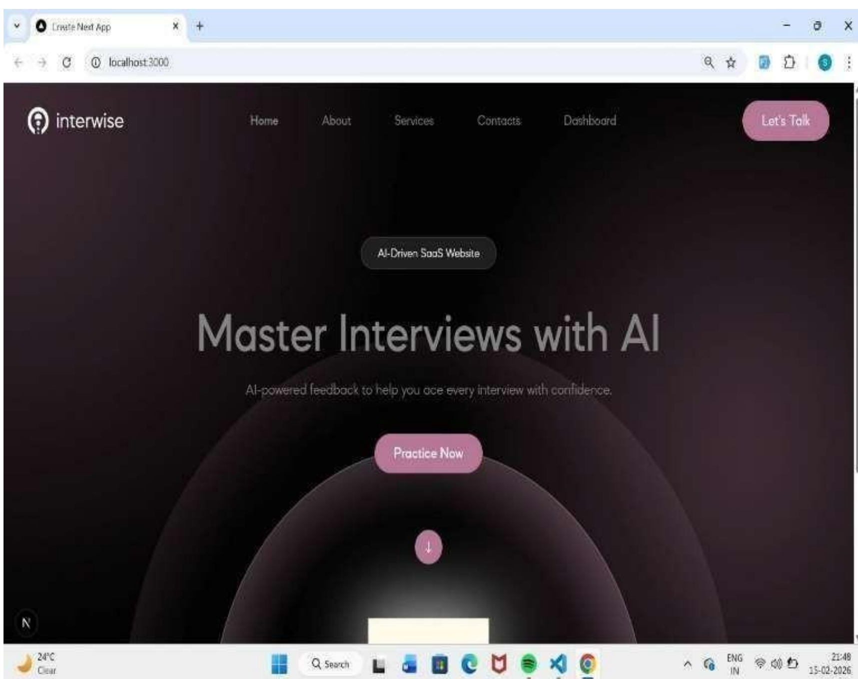


Fig. 2. The MIP interface

2. Data Labeling

After we had gathered all of the video interviews, we had six professionals with a minimum of ten years of expertise in the field rank them. Experts were given a five-point Likert scale to assess each video based on the interviewee's vocal quality, body language, facial expressions, interview substance, and whole performance. An individual's final annotated score was generated by summing the experts' judgements.

B. Feature Extraction

The software takes a number of elements out of uploaded video interviews for analysis at a later time. The extraction step covers: 1) visual extraction, 2) audio extraction, and 3) text analysis.

1. Visual Extraction

Since it is only possible to see the top half of a person while conducting a video interview, this section of the extraction concentrated on face analysis. Video frames were created by dividing each clip into one-second increments. In order to identify faces and extract (a) facial expressions and (b) head posture from each picture, this research used a face identification method.

a. Emotion: Eight different emotional states were identified by the facial recognition software: joy, surprise, anger, disgust, fear, disdain, and melancholy. All of the emotion qualities added up to a single score, which could be anything from zero to one. Finally, for the sake of visualization, the expressions were divided into three categories: positive (joy and surprise), neutral (other), and negative (anger, disgust, fear, disdain, and sorrow).

b. Head pose: The head stance is the manner in which the face is positioned in three-dimensional space. This characteristic is comprised of the roll, pitch, and yaw angles, which varied from -180 to 180 degrees (Fig. 3). In a head-tilting motion, the roll angle is on the x-axis, the pitch angle is on the y-axis, and the yaw angle is on the z-axis and specifies the nodding motion [21].

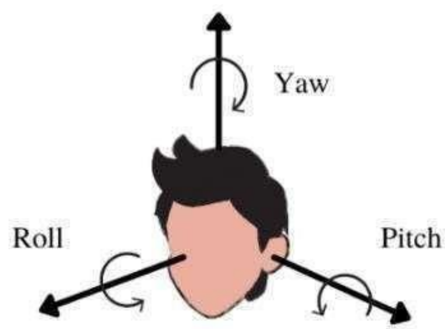


Fig. 3. Orientation of the head in terms of pitch, roll, and yaw movements.

2. Audio Extraction

This study collected three audio features: (a) speaking rate, (b) amplitude, (c) frequency. The audio was sliced into one-second intervals before performing feature extraction.

a. Speaking rate: The rate of speech is defined as the pace of speech minus the rate of speech during pauses [22]. Words uttered were identified in this investigation using speech-to-text technology developed for Traditional Chinese. Because they provide the appearance of dysfluency, pauses that are unfilled for more than three and a half seconds [23][24] were not included. The typical speaking speed of a native Chinese speaker is between 3.7 and 4.3 characters per second, according to many studies [25][26]. For that reason, this feature was split into three groups to help users better understand their own speaking speed: "slow" (0 to 2.5 cps), "medium" (2.5 to 4 cps), and "fast" (4 to 6 cps).

b. Amplitude: The relative intensity of sound waves, or amplitude, is quantified in decibels (dB) and is used to describe the volume of an audible signal [27]. The magnitude spectrum is derived from each audio interval of one second in order to compute amplitude. The largest magnitude (a) is taken and expressed in decibels (dB) by using the formula (1). A typical human voice, at a listening distance of about 70 dB, is audible to most people [27][28].

$$dB = 20 \log10(a) \quad (1)$$

c. Frequency: Hertz (Hz) is the standard unit of measurement for frequency, which is defined as "the number of waves that pass a given point in a given time period" [27]. Voiced speech typically has a fundamental frequency ranging from 100 to 300 Hz. We can determine the frequency of the audio interval by looking at the magnitude spectrum that was recovered before and finding the frequency that corresponds to the maximum magnitude.

3. Text Analysis

T His research classified interviewees' soft skill skills using two methods of personality trait analysis: 1) the DISC (Dominance, Influence, Steadiness, and Compliance) system, and 2) intrinsic qualities. As for the latter, there are a few things to keep in mind: (a) personality observability—you can tell if an interviewee has a dominant or recessive personality type just by watching how they talk and act during the interview; (b) ideal working style—you can learn a lot about an interviewee's interests and preferences just by looking at their voice and facial expressions, but you can also observe if they prefer working in a team or on their own.

C. Prediction Model

The following steps are implemented by processing the retrieved data. Following data extraction and preprocessing, a number of prediction models were used to examine the attributes. These models included 1) intrinsic qualities, 2) DISC model, 3) voice, and 4) emotion."

1. Emotion

Using the facial expressions of respondents, this characteristic encompasses eight distinct emotion factors. The distribution of each emotion as a percentage and the proportion of pleasant, neutral, and negative emotions per second were computed using the emotion variables. For this aspect, the researchers in this study used the Automatic Relevance Determination (ARD) prediction model, which yielded an ordinal "emotion score" from 1 to 5.

2. Pose

Using this characteristic, one may depict the position of the head in three dimensions. An analysis of the interviewee's head movement throughout the interview was conducted using a Gamma distribution model to train the head rotation data. The result was a "pose score" that ranged from 1 to 5.

3. Voice

A voice feature's components include the properties of pitch, volume, and frequency. An evaluation of amplitude and frequency was conducted each second, while speaking rate was averaged and categorized as "fast," "medium," or "slow." Then, all three variables were combined and trained using a linear regression model to forecast a "voice score" ranging from 1 to 5.

4. DISC Model

To determine the percentages of D, I, S, and C personality characteristics, interviewers gathered feature words that were associated with the DISC model and used them in calculations.

5. Intrinsic Traits

Through the use of natural language processing (NLP) technology, this research gathered and retrieved keywords from interview material. Using the interview transcripts of successful applicants from the past, we created a "semantic dictionary" to help with the selection process. After that, a "semantic score" was created by matching the related keywords taken from interviewees' material with the terms in the semantic dictionary.

In addition to the semantic score as the basis for determining intrinsic traits, this study proposed two models: 1) personality observability, and 2) ideal working style.

a. Personality observability: The interviewee's speech pace, frequency, and volume, as well as the change in "neutral" mood throughout the interview, were used in conjunction with text analysis—which includes semantic score and DISC score results—to provide a forecast for this measure.

These numbers play a significant role in the ARD regression model, a prediction model. As a general rule, a projected score between 1 and 5 indicates how "dominant" or obvious the interviewee's personality is; a score of 5 indicates the opposite.

b. Ideal working style: In text analysis, this function also integrated the semantic score with the findings of the DISC analysis. An individual's chosen method of operation may be significantly affected by subtle changes in their vocal intonation (volume, frequency, and pace of speech) and the angle at which they hold their head when answering questions.

The ARD regression model was also used as the prediction model for this characteristic. A score of 1 indicates that the interviewee is good at working alone, while a score of 5 indicates that they are more comfortable in a team setting.

RESEARCH RESULT

Mean Squared Error (MSE) and Mean Average Error (MAE) were computed for every model in order to assess their performance. The mean squared error (MSE) is calculated as (2) by averaging the squared difference between the actual data (yi) and the anticipated value ($\acute{y}$).MAE is the average of the absolute difference between the two values (3). Lower MSE and MAE values are preferred, which imply the estimated values are closer to the actual values.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \acute{y}_i)^2 \quad (2)$$

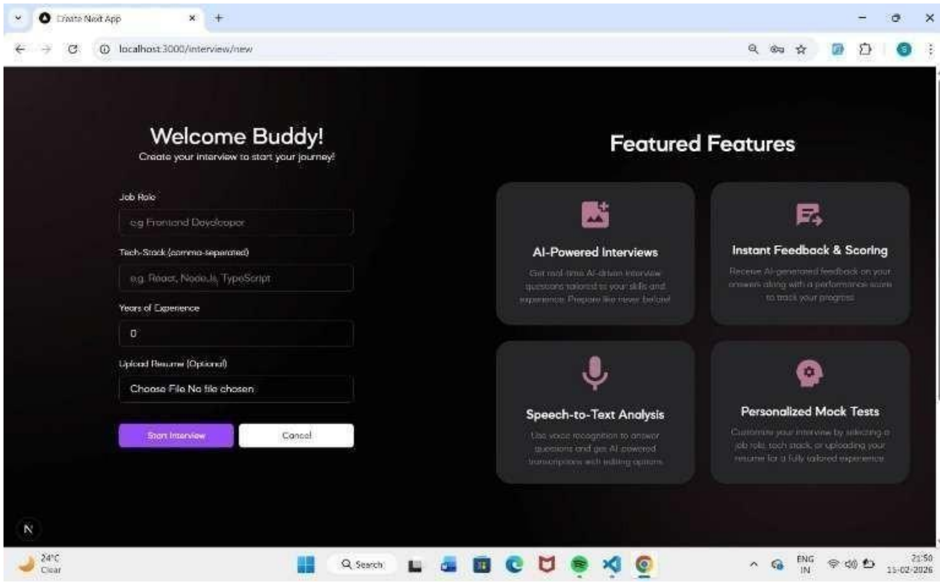
$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \acute{y}_i| \quad (3)$$

TABLEI. Analysis Methods and Prediction Results

Analysis	Model	Prediction results
Emotion	Automatic Relevance Determination (ARD)	MSE = 0.18 MAE = 0.34
Pose	Gamma distribution	MSE = 0.17 MAE = 0.34
Voice	Linear Regression	MSE = 0.21 MAE = 0.34
Personality observability	ARD	MSE = 0.11 MAE = 0.27
Ideal working style	ARD	MSE = 0.12 MAE = 0.25

Table 1 reports on the prediction results. The text-based models produced relatively lower MSE and MAE values compared to the other models. With an MSE of0.11, the personality observability model was the most accurate, followed by the ideal work style model with an MSE of 0.12. The personality observability model has an MAE of 0.27, while the optimum working style model has the lowest at 0.25. The most significant error (MSE) was 0.21 for the speech analysis, followed by 0.18 for the emotion model and 0.17 for the posture model. A mean absolute error of 0.34 was recorded by the models for emotion, posture, and voice.

To help evaluate the interviewee's performance, the algorithms covered earlier in this section generated quantifiable predicted scores. Upon completion of the interview, the platform immediately generates an analytical report displaying these results. The findings of the analysis of voice, emotion, and personality characteristics, together with the overall performance ratings, are shown in the four columns of the MIP report (Fig. 4). Here we can see the results of the various analyses: the voice analysis shows the intonation change per second, speech rate (words/second), and volume change per second. The emotion analysis shows the overall distribution of 8 facial expressions and the emotion change per second. The personality traits analysis shows the DISC traits proportion as well as the results of the ideal working style model and personality observability. Finally, the overall performance score includes the voice, pose, emotion, and semantic scores, as well as an aggregated final score.



IV. CONCLUSIONS AND FUTURE WORKS

This research proposes a mock-interview platform that uses visual, audio, and textual elements to assess applicants' personality characteristics and performance during interviews. Applicants may be evaluated objectively and important information that people miss can be detected by the platform. The models that were used in this investigation yielded good results. It is challenging to undertake a comparison study on the performance outcomes due to the absence of comparable items in the market. We think there's room to improve the precision of emotion, posture, and voice models. To create a more complete model, future research should include characteristics like eye gaze direction and verbal fluency. To further improve the model's accuracy, weighting should be implemented according to the requirements of businesses.

Human resources managers and supervisors may speed through the application screening process with the help of the platform's analysis report. As an added bonus, interviewees may utilize this platform to practice interviews on their own. Candidates may have a better grasp of the interview process and hone their abilities using this platform. The platform's current use case has room to grow into other areas; for instance, schools might use it to better prepare their incoming freshmen for college interviews or job interviews.

The platform only offers a broad overview of the characteristics at the moment in its analysis reports. Depending on the needs of HR or supervisors, the platform might be adjusted in the future to provide various analytical reports. It would be beneficial to tailor the platform's interview questions to each specific job opening, taking into account the wide range of businesses offering employment opportunities. It would be possible to create a bank of interview questions that are tailored to each role. To get a more complete picture, you might also give each question a certain weight.

REFERENCES

[1]. J. Smith, “Thomas Edison conducted the first job interview in 1921 - here's how they've evolved since,” Business Insider, May 21, 2015. [Online]. Available: https://www.businessinsider.com/evolution-of-the-job-interview-2015-5

[2]. T. Ackermans, “The honest truth about your job search,” Portl, May 21, 2021. [Online]. Available: https://portl.nl/the-honest-truth-about-your-job-search/

[3]. “27 Interview Statistics: All You Need to Know in 2021,” What To Become, August 11, 2021. [Online]. Available: https://whattobecome.com/blog/interview-statistics

[4]. “Three Approaches to Effectively Manage Virtual Interviews,” Gartner Research, October 29, 2020. [Online]. Available: https://www.gartner.com/en/documents/3992422

[5]. M. Lazic, “30 mind-blowing interview statistics to get you going in 2021,” LegalJobs, February 1, 2021. [Online]. Available: https://legaljobs.io/blog/interview-statistics/

[6]. E. Utami, A. D. Hartanto, S. Adi, I. Oyong, and S. Raharjo, “Profiling analysis of disc personality traits based on Twitter posts in Bahasa Indonesia,” Journal of King Saud University - Computer and Information Sciences, 2019.

[7]. L. Zhang, Z. X. Chen, and B. Yang, “Personality Analysis and Prediction of Social Network Users,” Chinese Journal of Computers, vol. 37, no. 8, August 2014.

[8]. Y. C. Yao, “Artificial Intelligence simulate interviewer’s prediction of job applicants’ communication skill and big five personality traits,” M.S. Thesis, Dept. of Technology Application and Human Resource Development, National Taiwan Normal Univ., Taipei, 2020.

[9]. Y. C. Chou and H. Y. Yu, “Based on the application of AI technology in resume analysis and job recommendation,” 2020 IEEE International Conference on Computational Electromagnetics (ICCEM), 2020, pp. 291-296.

[10]. Frequently asked questions, HireVue. [Online]. Available: https://www.hirevue.com/candidates/faq

[11]. D. Harwell, “A face-scanning algorithm increasingly decides whether you deserve the job,” The Washington Post, November 6, 2019. [Online]. Available: https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/

[12]. “What are AI interviews: Understanding enterprise recruitment trend,” Aurora Cloud, June 18, 2020. [Online]. Available: https://www.aurora.com.tw/cloud/column/0k125416353273783414.

[13]. “Comparison of AI interview systems: effectively recruit talents with Lasso,” MAYO Human Capital. [Online]. Available: https://www.mayohr.com/tw/blog/detail/Lasso-AI-interview.

[14]. A. Mehrabian and S. R. Ferris, “Inference of attitudes from nonverbal communication in two channels,” Journal of Consulting Psychology, vol. 31, no. 3, pp. 248-252, 1967.

[15]. Y. Adepu, V. R. Boga and S. U, "Interviewee Performance Analyzer Using Facial Emotion Recognition and Speech Fluency Recognition," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-5.

[16]. L. Chen, R. Zhao, C. W. Leong, B. Lehman, G. Feng, and M. E. Hoque, “Automated video interview judgment on a large-sized corpus collected online,” 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), 2017.

[17]. I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, “Automated Analysis and prediction of Job Interview Performance,” IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 191–204, 2018.

[18]. V. J. Delima, “Impact of Personality Traits on Employees' Job Performance in Batticaloa Teaching Hospital,” Iconic Research and Engineering Journals, vol. 2, no. 12, June 2019.

[19]. B. L. L. A. Balasuriya and G. D. N. Perera, “The Impact of Personality on Job Satisfaction: A Study of Executive Employees in Selected Private Hospitals in Colombo East, Sri Lanka,” IJRDO – Journal of Business Management, vol. 2, no. 12, pp. 7-15, December 2016.

[20]. H.-Y. Suen, K.-E. Hung, and C.-L. Lin, “Intelligent video interview agent used to predict communication skill and perceived personality traits,” Human-centric Computing and Information Sciences, vol. 10, no. 1, 2020.

[21]. E. N. Arcoverde Neto, R. M. Duarte, R. M. Barreto, J. P. Magalhães, C. M. Bastos, T. I. Ren, and G. D. C. Cavalcanti, “Enhanced real-time head pose estimation system for mobile device,” Integrated Computer-Aided Engineering, vol. 21, no. 3, pp. 281–293, 2014.

[22]. E. Jacewicz, R. A. Fox, C. O'Neill, and J. Salmons, “Articulation rate across dialect, age, and gender,” Language Variation and Change, vol. 21, no. 2, pp. 233–256, 2009.

[23]. L. F. Huang and T. Graf, “Speech Rate and Pausing in English: Comparing learners at different levels of proficiency with native speakers,” Taiwan Journal of TESOL, vol. 17, no. 1, pp. 57-86, 2020.

[24]. E. Campione and J. Veronis, “A Large-Scale Multilingual Study of Silent Pause Duration,” in Speech Prosody 2002, pp. 199-202, 2002.

[25]. A. H. S. Chan and P. S. K. Lee, “Intelligibility and preferred rate of Chinese speaking,” International Journal of Industrial Ergonomics, vol. 35, no. 3, pp. 217–228, 2005.

[26]. C. H. Wu, “Filled Pauses in L2 Chinese: A Comparison of Native and Non-Native Speakers,” Proceedings of the 20th North American Conference on Chinese Linguistics, vol. 1, pp. 213-227, 2008.

[27]. “Introduction to Psychology. Module 4: Sensation and Perception, Waves and Wavelengths,” Lumen. [Online].Available: https://courses.lumenlearning.com/atd-bhcc-intropsych/chapter/waves- and-wavelengths/.

[28]. “Facts about speech intelligibility: Human voice frequency range,” DPA Microphones, Mar 3,2021. [Online]. Available: https://www.dpamicrophones.com/mic-university/facts-about-speech-intelligibility.