Real Time Mock Interview using Deep Learning

Rohan Patil [1], Akash Butte [2], Sahil Temgire [3], Varun Nanekar [4], Asst. Prof. Shivganga Gavhane [5]

[1, 2, 3, 4] Student, Department of Computer Engineering, DYPIEMR Akurdi, Pune, India

[5] Professor, Department of Computer Engineering, DYPIEMR Akurdi, Pune, India

Abstract: **Real Time Mock Interview Using Deep Learning system is a web application helpful for users to practice for interviews. Nowadays many companies are conducting interviews virtually through online mode. So, this is the need of the day to develop a system where users can practice for these online interviews. This system will help candidates to practice for mock interviews by facing mock interviews. It also provides feedback including facial preference, head nodding, reaction time, speaking rate and volume to let users know their own performance within the mock interview. The system provides speech-to-text conversion for checking grammar in the candidates reply and suggests required corrections. Results are given in a graphical format by using these two or more interviews can be compared to track the progress of the candidates and corrective action will be taken in order to give better performance in the next interviews.**

Keywords: *Face expression recognition, convolutional neural network, Speech-to-text, Deep learning, Grammar specification language (GSL).*

I. INTRODUCTION

Recently college graduates often have the chance to participate within the interview once they attempt to pursue further studies or find employment. So as to master all possible questions within the interview, the simplest way is to understand what sorts of questions could also be asked and practice responding to questions. Generally, college students rarely have the chance to practice interview during school. So as to extend opportunities for people to practice social skills, like admission interview and employment interview, many scholars engaged within the design and development of social skill training systems Job interviews are employed by the potential future employer as a way to work out whether the interviewee is fitted to the company's needs. To form an assessment, interviewers heavily rely on social cues, i.e. actions, conscious or unconscious, of the interviewee that have a selected meaning during a social context, like employment interview. During this paper an approach is presented to employment interview simulation environment which uses a social virtual character as a recruiter and signal processing techniques to enable the virtual character to react and adapt to the user's behaviour and emotions. The aim of this simulation is to assist youngsters improve social skills which are pertinent to job interviews. The proposed system features a real-time social cue recognition system, a dialog/scenario manager, a behaviour manager and a 3D rendering environment. The next section offers a quick review of the interdisciplinary literature. This system provides some feedbacks including facial preference, head nodding, reaction time, speaking rate and volume to let users know their own performance within the mock interview. Using speech-to-text system will check the grammar. The system will provide the result in a graphical format. The result of two or more interviews can be compared to track the progress of the candidates.

II. LITERATURE SURVEY:

This study proposes an approach to dialog state tracking and action selection supported deep learning methods. First, the interview corpus during this study is collected from 12 participants, and is annotated with dialog states and actions. Next, a long-short term memory and a man-made neural network are employed to predict dialog states and therefore the Deep RL is adopted to find out the relation between dialog states and actions. Finally, the chosen action is employed to get the interview question for interview practice. to gauge the proposed method in action selection, an interview coaching system is made. Experimental results show the effectiveness of the proposed method for dialog state tracking and action selection. In this study, an interview coaching system is proposed and constructed for dialog state tracking and action selection. LSTM and ANN are employed to predict dialog states and the deep RL is used to learn the relation between dialog states and actions. Finally, the anticipated dialog state and action pair are wont to generate an interview question. For performance evaluation on the proposed method in dialog state tracking and action selection, an interview coaching system was constructed, AND AN encouraging result was obtained for dialog state tracking, action selection and interview question generation [2].

It is documented that syntactic constraints, when applied to speech recognition, greatly improve accuracy. However, until recently, constructing an efficient grammar specification to be used by a connected word speech recognizer was performed by hand and has been a tedious, time-consuming task susceptible to error. For this reason, very large grammars haven't appeared. We describe a compiler for constructing optimized syntactic di- graphs from easily written grammar specifications. These are written during a language called grammar specification language (GSL). The compiler features a pre-processing (macro expansion) phase, a parse phase, graph code generation and compilation phases, and three optimization phases. Digraphs also can be linked together by a graph linker to make larger digraphs. Language complexity is analysed during a statistics phase. Heretofore, computer generated digraphs were often crammed with redundancies. Larger graphs were constructed and optimized by hand so as to realize the specified efficiency. We demonstrate that the optimization phase yields graphs with even

greater efficiency than previously achieved by hand. We also discuss some preliminary speech recognition results of applying these techniques to intermediate and enormous graphs. With the introduction of those tools it is now possible to supply a speech recognition user with the power to define new task grammars within the field. GSL has been employed by several untutored users with good success. Experience with GSL indicates that it's a viable medium for quickly and accurately defining grammars to be used in connected speech recognition systems [3].

With the development of artificial intelligence (AI), the automatic analysis of video interviews to recognize individual personality traits has become an active area of research and has applications in personality computing, human-computer interaction, and psychological assessment. Advances in computer vision and pattern recognition based on deep learning (DL) techniques have led to the establishment of convolutional neural network (CNN) models that can successfully recognize human nonverbal cues and attribute their personality traits with the utilization of a camera. during this study, an end-to-end AI interviewing system was developed using asynchronous video interview (AVI) processing and a Tensor Flow AI engine to perform automatic personality recognition (APR) supported the features extracted from the AVIs and therefore the true personality scores from the facial expressions and self-reported questionnaires of 120 real job applicants. The experimental results show that our AI-based interview agent can successfully recognize the "big five" traits of an interviewee at an accuracy between 90.9% and 97.4%. Our experiment also indicates that although the machine learning was conducted without large-scale data, the semi supervised DL approach performed surprisingly well with reference to automatic personality recognition despite the lack of labor-intensive manual annotation and labeling. The AI-based interview agent can supplement or replace existing self-reported personality inventory methods that job applicants may distort to realize socially desirable effects [4].

To avoid the complex process of explicit feature extraction in traditional countenance recognition, a face recognition method supported a convolutional neural network (CNN) and a picture edge detection is proposed. Firstly, the countenance image is normalized, and therefore the fringe of each layer of the image is extracted within the convolution process. The extracted edge information is superimposed on each feature image to preserve the sting structure information of the feel image. Then, the dimensionality reduction of the extracted implicit features is processed by the utmost pooling method. Finally, the expression of the test sample image is assessed and recognized by employing a Soft max classifier. To verify the robustness of this method for countenance recognition under a posh background, a simulation experiment is meant by scientifically mixing the Fer-2013 countenance database with the LFW data set. The experimental results show that the proposed algorithm are able to do a mean recognition rate of 88.56% with fewer iterations, and therefore the training speed on the training set is about 1.5 times faster than that on the contrast algorithm [5].

To explore human emotions, during this paper, we design and build a Multi-Modal Physiological Emotion Database (MPED), which collects four modal physiological signals, i.e., electroencephalogram (EEG), galvanic skin response (GSR), respiration (RSP) and electrocardiogram (ECG). To alleviate the influence of culture dependent elicitation materials and evoke desired human emotions, we specifically collect an emotion elicitation material database selected from quite 1500 video clips. By considerable amount of strict man-made labelling, we elaborately choose 28 videos as standardised elicitation samples, which are assessed by psychological methods. The physiological signals of participants were synchronously recorded once they watched these standardised video clips that described six discrete emotions and neutral emotion. With three kinds of classification protocols, different feature extraction methods and classifiers (SVM and KNN) were used to recognize the physiological responses of varied emotions, which presented the baseline results. Simultaneously, we present a totally unique attention-LSTM (A-LSTM) which strengthens the effectiveness of useful sequences to extract more discriminative features. Additionally, correlations between the EEG signals and thus the participants' ratings are investigated. The database has been made publicly available to encourage other researchers to use it to gauge their own emotion estimation methods

Emotion detection and recognition from text could also be a recent essential research area in tongue Processing (NLP) which may reveal some valuable input to a selection of purposes. Nowadays, writings take many sorts of social media posts, micro-blogs, news articles, customer review, etc., and thus the content of those short-texts are often a useful resource for text mining to urge an unhide various aspects, including emotions. The previously presented models mainly adopted word embedding vectors that represent rich semantic/syntactic information and other people models cannot capture the emotional relationship between words. Recently, some emotional word embeddings are proposed but it requires semantic and syntactic information the other way around. to affect this issue, we proposed a completely unique neural specification, called SENN (Semantic-Emotion Neural Network) which can utilize both semantic/syntactic and emotional information by adopting pre-trained word representations. SENN model has mainly two sub-networks, the first sub-network uses bidirectional Long-Short Term Memory (BiLSTM) to capture contextual information and focuses on semantic relationship, the second sub-network uses the convolutional neural network (CNN) to extract emotional features and focuses on the emotional relationship between words from the text. We conducted a comprehensive performance evaluation for the proposed model using standard real-world datasets. We adopted the notion of Ekman's six basic emotions. The experimental results show that the proposed model achieves a significantly superior quality of emotion recognition with various state-of-the-art approaches and further are often improved by other emotional word embedding [7].

III. PROPOSED SYSTEM

For our interactive scenario we rely on a software frame-work that supports a fine grained multimodal behaviour control for virtual characters, In this environment, the virtual character plays the role of a recruiter which reacts and adapts to the user's behaviour thanks to a component for the automatic recognition of social cues (conscious or unconscious behavioural patterns). The social cues pertinent to job interviews have been identified using a knowledge elicitation study with real job seekers. Finally, we present two user studies to investigate the feasibility of the proposed approach as well as the impact of such a system on users.
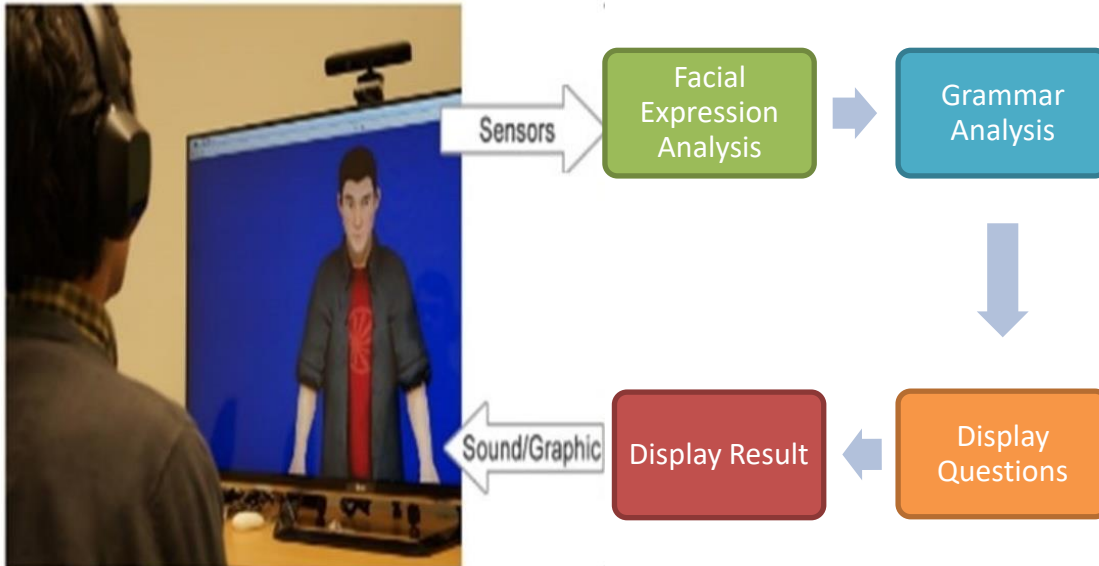
Fig 1. System Design
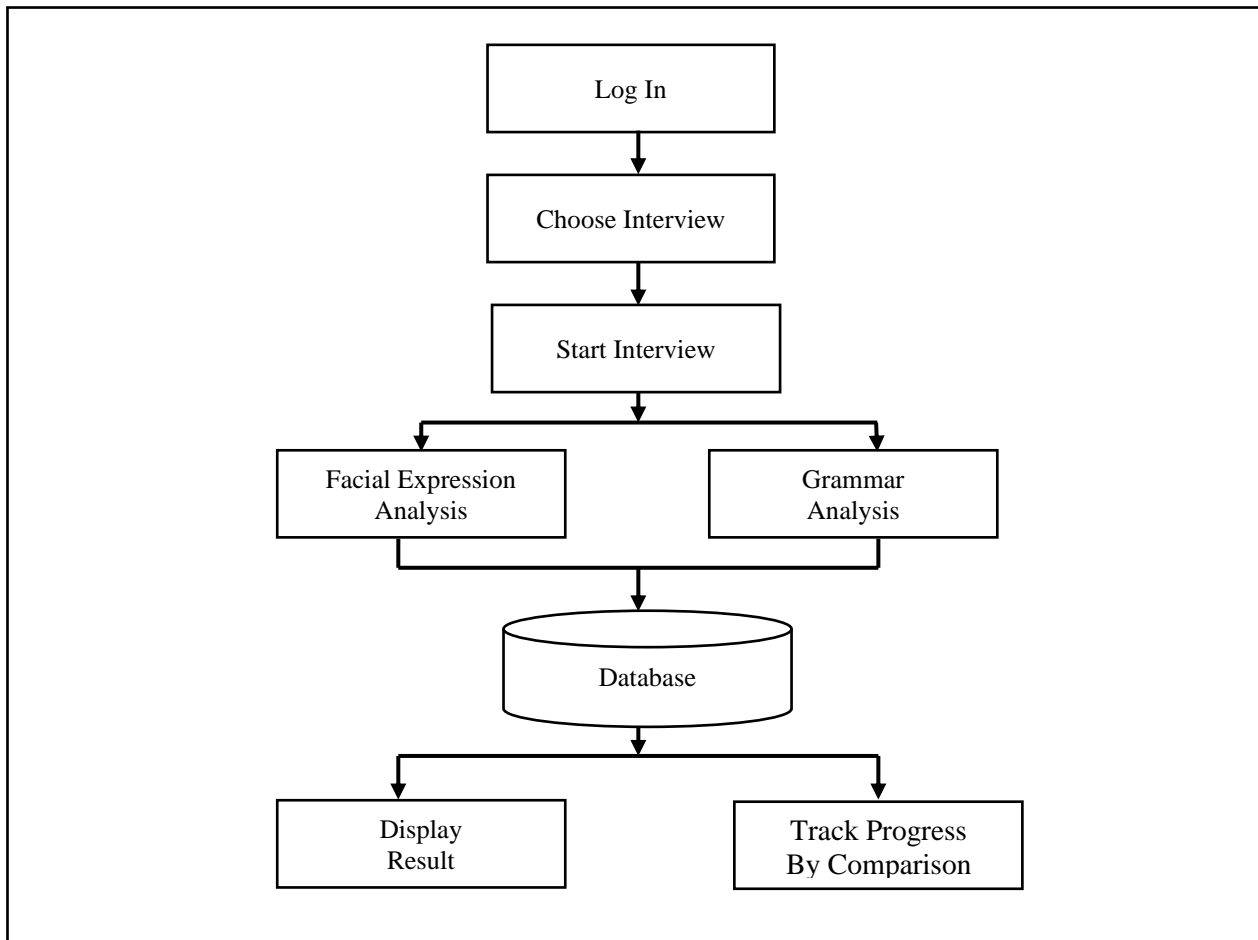
Fig 2. Architecture Diagram

Log In – First the user has to sign up in the system. After that user will get a username and password. Using the credentials, user can login in the system.

Choose Interview – After successfully logging in, user will have to choose interview of his choice based upon the interviews present in the system. User can choose interviews to track his performance and progress which will be saved in the database while comparing the results.

Start Interview – Once the user chooses the interview of his preference, after that the interview will start, while the interview proceeds users progress and performance will be saved in the database of the system. User's progress and performance saved in the database will be further used to compare different parameters at the end while displaying the result.

Facial Expression Analysis – When the user is giving the interview with the help of the webcam user's facial expression will be analysed. When the user chooses to start interview the OpenCV runs in background and starts to record the video of the interview. This facial expression will be analysed based on the dataset imported in the system and this data will be stored in the database. This proposed system will help users who are nervous or anxious while giving the interviews. The interviewers keep a keen attention on the expressions of the candidates they are interviewing because many candidates are rejected because they are not confident while giving the interviews. Therefore, facial expression analysis is required or essential so that the interviewee can improve his performance in front of the panel interviewing the user by taking mock interviews which will help the interviewee while giving actual company recruitment interviews.

Grammar Analysis – When the user will be giving the interview, whatever user speaks will be converted into text. This conversion is essential to keep an eye on user's grammatical mistakes. This analysis will help the user improve his vocabulary for actual interviews. Along with confidence of the interviewee, interviewers also keep an eye on the vocabulary of the candidate. Therefore, this analysis will be saved in the database which will be used while displaying result to make user aware of his weaknesses in vocabulary. Hence the user can work on the factors which are affecting his performance in actual interviews.
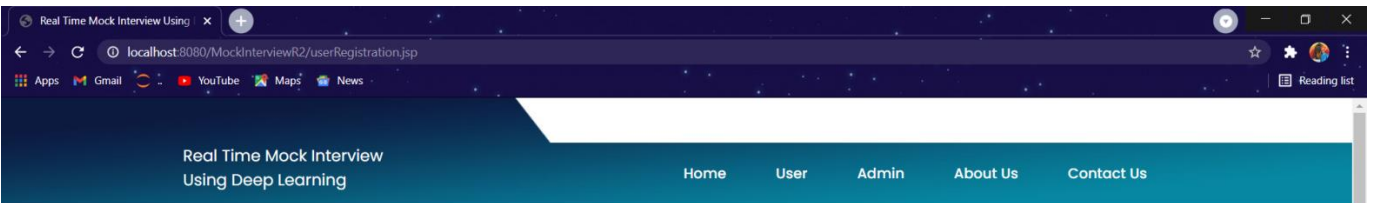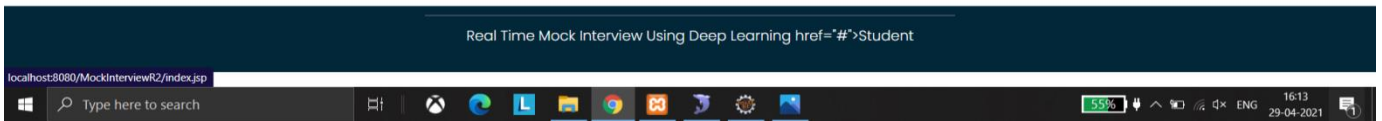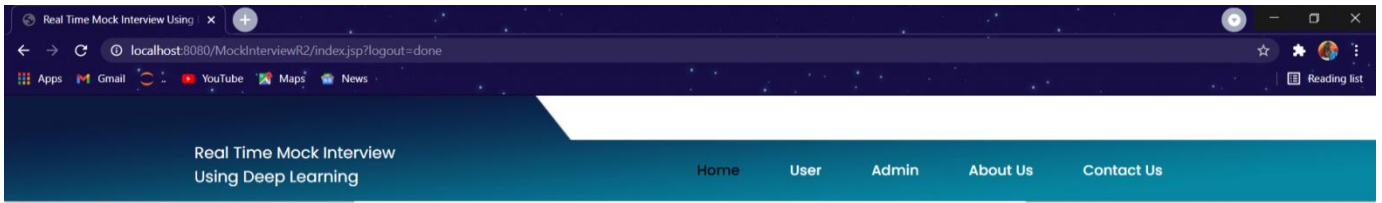
Database – Database is an important factor of this proposed system because user's progress and performance are saved in the database. This saved data of user will help to improve performance every time he logs in the system to practice for actual interviews. Also, user's data like user's preferences, personal information will be also saved in the database.

Display Result – After user completes the interviews based on performance and progress result will be displayed. This result will help user to improve on the factors in which the result is bad. Result will be displayed based on data analysis visualization. User's facial expression analysis and grammar analysis will be displayed in the result. As the facial expressions and grammatical errors will be converted to some dataset, the data will be then analysed and predicted. Both factors will be considered and result will be displayed in graphical format using data visualization tools. The result of two or more interviews can be compared to track the progress of the candidates. The result of users performance in particular interview will also depend on his facial preference, head nodding, reaction time and speaking rate.

A. *FEATURES OF THE SYSTEM*
- This system provides time efficient and very effective candidate selection mechanism.
- It is highly customizable as employer can specify their criteria along with importance level.
- It is easy for user as they just need to upload their resumes on portal.
- No form filling is required.
- Automatic Email notification to candidate / employers can be possible.
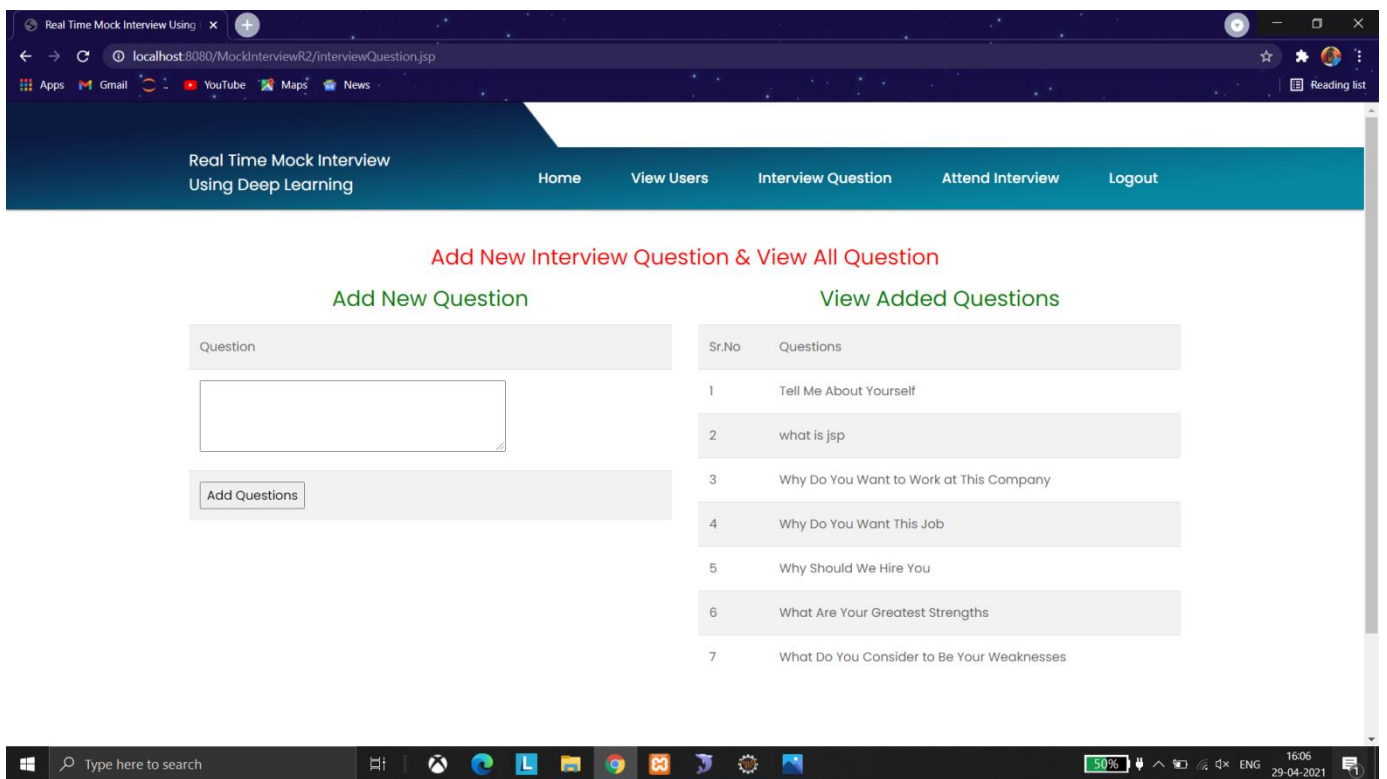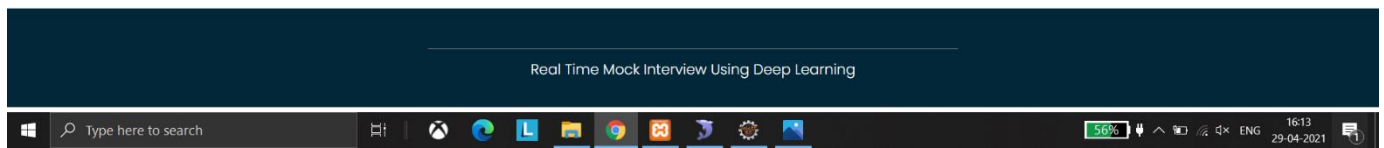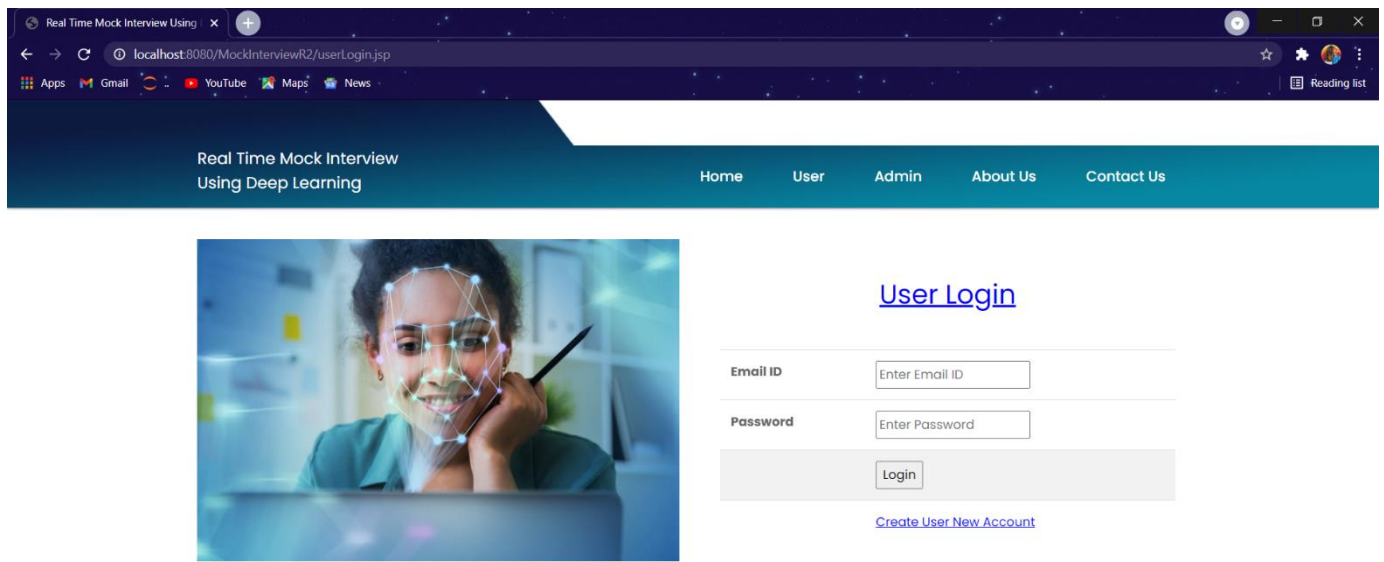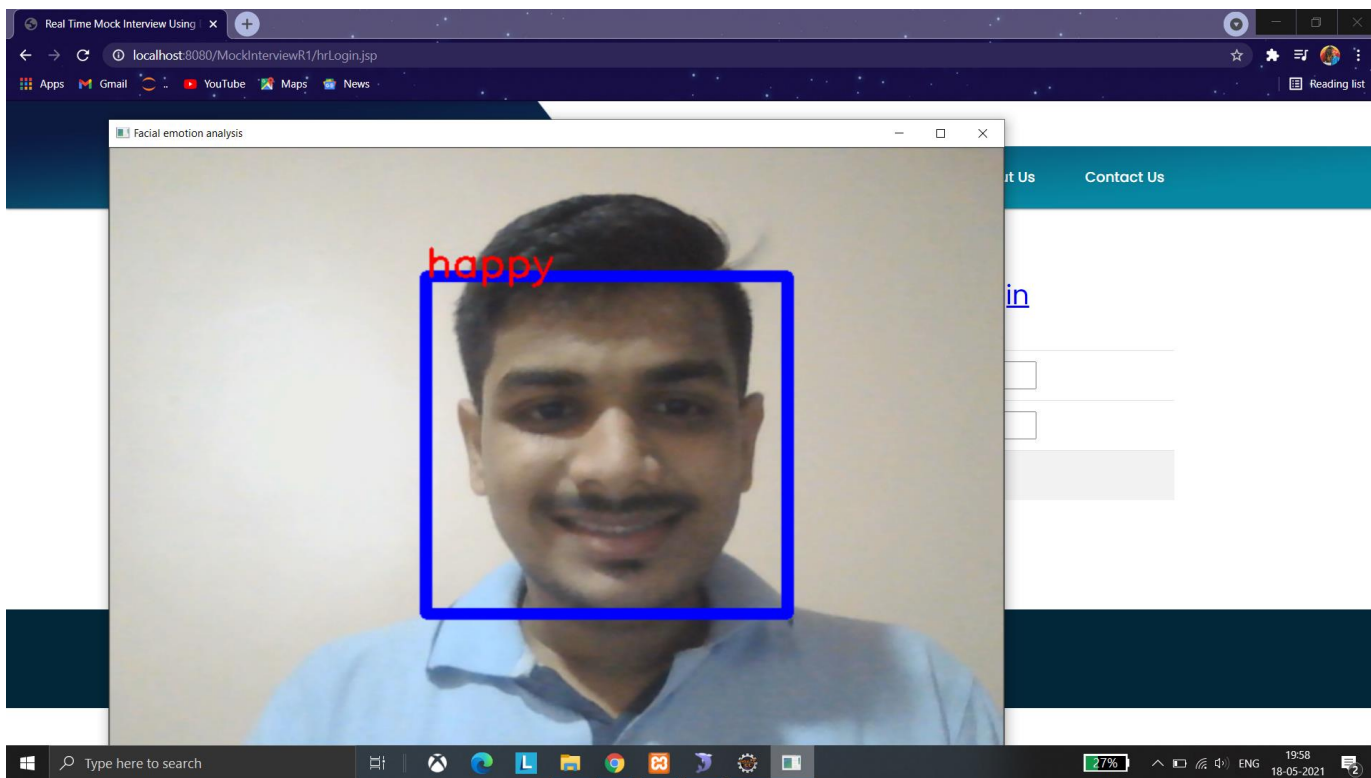
IV. EXPERIMENTAL RESULTS

V. CONCLUSION

In this paper, we presented an approach that enhances a virtual agent by the ability to interpret and respond to social cues of users participating in a simulated job interview. In order to achieve seamless credible interaction, our system automatically recognizes the user's social cues in real time. Based on these, the virtual recruiter reacts and adapts to the user's behaviour. Furthermore, the interaction with the virtual agent can be recorded and presented to the user to enhance the learning effect, for example, by identifying critical incidents during the simulated interview. The scenario manager was used to model the virtual recruiter's interactive behaviour allowing the character to react to various social users recognized by the social cue recognition module. More precisely, we modelled mirroring and turn taking behaviour. Despite several reported problems, such as the realism of the character's appearance, all participants' reactions were mainly positive saying they would use such a system to train for real job interviews.

VI. REFERENCES

[1] Ming-Hsiang Su, Kun-Yi Huang, Tsung-Hsien Yang, Kuan-Jung Lai and Chung-Hsien Wu " DIALOG STATE TRACKING AND ACTION SELECTION USING DEEP LEARNING MECHANISM FOR INTERVIEW COACHING".

[2] "Michael K. Brown and Jay G. Wilpon"A Grammar Compiler for Connected Recognition.

[3] Hung-Yue Suen1, Kuo-En Hung1, and Chien-Liang Lin "Tensor Flow-based Automatic Personality Recognition Used in Asynchronous Video Interviews"

[4] Hongli Zhang, Alireza Jolfaei, And Mamoun Alazab "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing".

[5] Tengfe Song, Wenming Zheng, Cheng Lu, Yuanzong, Xilei Zhang1and Zhen Cui "MPED: A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition" IEEE Access, vol. 7, pp. 12177–12191, 2019.

[6] Erdenebileg Batbaatar, Meijing Li, And Keun Ho Ryu "Semantic-Emotion Neural Network for Emotion Recognition from Text" IEEE Access, vol. 7, pp. 111866–111878,2019.

[7] Michael K. Brown and Jay G. Wilpon "A Grammar Compiler for Connected Recognition Speech" IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 39. NO. 1, JANUARY 1991

[8] R. MacDonald, "Disconnected youth? social exclusion, the underclassand economic marginality,"Social Work and Society, vol. 6, no. 2, pp.236–247, 2008.

[9] T. Hammer, "Mental health and social exclusion among unemployedyouth in scandinavia. a comparative study,"Intl. Journal of SocialWelfare, vol. 9, no. 1, pp. 53–63, 2000. [Online]. Available:http://dx.doi.org/10.1111/1468-2397.00108.

[10] R. D. Arvey and J. E. Campion, "The employment interview: Asummary and review of recent research,"Personnel Psychology,vol. 35, no. 2, pp. 281–322, 1982. [Online]. Available: http://dx.doi.org/10.1111/j.1744-6570.1982.tb02197.x

[11] J. Curhan and A. Pentland, "Thin slices of negotiation: predictingoutcomes from conversational dynamics within the first 5 minutes,"pp. 802–811, 2007.

[12] N. Bianchi-Berthouze, "Understanding the role of body movement inplayer engagement,"HumanComputer Interaction, vol. 28, no. 1, pp.40–75, 2013. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07370024.2012.688468

[13] J. Greene and B. Burleson,Handbook of Communication and SocialInteraction Skills, ser. LEA's Communication Series. L. ErlbaumAssociates, 2003. [Online]. Available: http://books.google.de/books?id=h3i-refbDKwC

[14] H. Prendinger and M. Ishizuka, "The empathic companion: Acharacter-based interface that addresses users' affective states,"Applied Artificial Intelligence,   vol.   19,   no.   3-4,   pp.   267–285,   2005.[Online].   Available: http://www.ingentaconnect.com/content/tandf/uaai/2005/00000019/F0020003/art00004

[15] B. Endrass, E. Andŕe, M. Rehm, and Y. Nakano, "Investigating culture-related aspects of behavior for virtual characters,"Autonomous Agentsand Multi-Agent Systems, 2013