2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)

Job Interview Training System using Multimodal Behavior Analysis

Nao Takeuchi
Graduate School of Information Science and Technology
Osaka Institute of Technology
Osaka, Japan
m1m21a29@st.oit.ac.jp

Tomoko Koda
Faculty of Information Science and Technology
Osaka Institute of Technology
Osaka, Japan
tomoko.koda@oit.ac.jp

***Abstract*—The paper introduces our system that recognizes the nonverbal behaviors of an interviewee, namely gaze, facial expression, and posture using a Tobii eye tracker and cameras. The system compares the recognition results with those of models of exemplary nonverbal behaviors of an interviewee and highlights the behaviors that need improvement while playing back the interview recording. The development goal for our system was to construct an inexpensive and easy-to-use system using commercially available HWs, open-source code, and a CG agent that would provide feedback to the interviewee. The results of the initial evaluation of the system indicate that improvements in the recognition accuracy of nonverbal behaviors and the quality of the interaction with the CG agent are needed.**

***Index Terms*—multi-modal interaction, gaze recognition, posture recognition, facial expression recognition, job interview training, nonverbal behavior, CG agent**

1. INTRODUCTION

Interview training can help students acquire skills by experiencing the content and flow of job interviews and can increase their confidence in their search for employment. However, interview training has been limited due to the number of interviewers and time available for interviewing [1]. Moreover, the impact of COVID-19 has made interview training more difficult to conduct. This suggests that there is an increasing need for a system that allows students to train for job interviews independently.

Visual nonverbal behavior during a dialog accounts for 55% of all the information conveyed [2]. Washburn et al. pointed out that the outcome of an interview is affected more by the nonverbal behaviors of an interviewee than their verbal behaviors [3]. Moreover, Arvey et al. noted that nonverbal behaviors such as gaze, body movements, and tone of voice greatly influence the interviewee's evaluation [4]. These studies show that the use of nonverbal behavior and its impact on job interview success has been a major focus in research.

In recent years, social signal processing techniques using multimodal information have been used for dialog analysis [5] and have been applied to AI-based interview recruitment systems [6, 7, 8, 9] and interview training systems [10, 11, 12, 13]. Specifically, there are those that visualize the information of the nonverbal behavior and provide feedback during or after the interview [14, 15, 16, 17], and those that change the behavior of the interviewer, i.e., the CG agent [18, 19, 20]. However, most of the research conducted in the field of social signal processing focused on the recognition of emotions based on speech and facial expressions and paid less attention to posture recognition. In addition, some studies [6, 7, 9] proposed high-end computer systems that are not affordable for general users.

It has also been reported that practicing interviewing with a CG agents is more effective in improving skills compared to using books and videos on job interviews [15, 17]. Other interview practice systems using CG agents have been found to elicit self-disclosure [21, 22]. [22] has shown using a CG agents increase user's self-disclosure and feelings of rapport, self-efficacy, and trust.

Consequently, the purpose of this study was to develop an interview training system specializing in the recognition of three types of nonverbal behaviors, namely, gaze, facial expression, and posture, and to utilize a CG agent that would provide feedback on the appropriateness of the nonverbal behaviors of interviewees. We expect this system allows people to practice their interview skills by themselves.

2. JOB INTERVIEW TRAINING SYSTEM

2.1. System Overview

The system was developed using Unity, FaceAPI [23], OpenPose [24], TobiiEyeTracker4C [25], and a webcam. This system consists of two phases: a mock interview phase, and a feedback phase. During the mock interview phase the interview was video-recorded from a front-left angle in order to obtain the nonverbal behaviors of the interviewee, including gaze, facial expression, and posture. The captured video was analyzed by the following procedures (see 2.2) and played back in the feedback phase. The system paused the video where feedbacks were needed and the CG agent provided feedbacks on any points for improvement.

2.2. Detection of Nonverbal Behaviors

2.2.1 Detectable Nonverbal Behaviors

The nonverbal behaviors were acquired at 1-second intervals for gaze and 3-second intervals for facial expression and posture. From the acquired behaviors, the following information was obtained: gaze rate, user's gaze moving off the interviewer's face for more than five seconds, number of times the gaze point moved to the upper right or upper left, level of smile or straight face, six facial expressions (anger, contempt, disgust, fear, sadness, surprise), posture (forward and backward leaning), legs open, legs opening gradually, shake of the neck, and protrusion of the elbows.

2.2.2 Gaze Detection Method

We used a collision-detection method in order to detect inappropriate gazes. In order to determine when the interviewee's gaze moved off of the interviewer's face, we preset an area-of-interest (AOI) on the interviewer's entire

A part of this work was supported by Grant-in-Aid for Scientific Research "KAKENHI (C) 20K11926".

978-1-6654-0021-3/21/$31.00 ©2021 IEEE

face. The system determined inappropriate gaze when the interviewees gazing point moved out of the AOI for five seconds. The gaze rate was calculated by dividing the number of frames in which the interviewee was looking at the interviewer's face by the total number of frames and displaying it as a percentage. The number of times the gazing point moved to the upper right or upper left was determined by setting up another AOI in the upper right and upper left areas of the screen(next to the interviewer). The system determined inappropriate gaze if the gazing point entered these areas more than 10 times, it was counted. Eventually, these metrics were used to determine the feedback given by the CG agent during the feedback phase.

2.2.3 Facial Expression Detection Method

For inappropriate facial expression detections, we used “smile” and “emotion” recognition scores from FaceAPI. The smile and straight face scores were set at 0 and 1, respectively. A smile was determined as a result when a score of 0.5 or more was detected, and a straight face was determined as the result when 0 was detected three times in a row.

2.2.4 Posture Detection Method

Inappropriate postures were detected by comparing a correct posture model with the posture of the interviewee. A correct posture model was created for each gender using OpenPose under the guidance of the Employment Department of our university. For example, whether an interviewee was leaning forward and backward was judged when there was a difference of more than 20 degrees between the model and the interviewee.

2.3. Feedback Algorithm

The detected information on inappropriate nonverbal behaviors was stored in the gaze, facial expression, and posture arrays each second and then compiled into a single array using weighted prioritization. The weight was set in the order of gaze, facial expression, and posture, based on the order of importance during the interview. This also helped to avoid duplication of multiple detections in the same number of seconds and biased results that pointed to the same type of feedback. The CG agent would refer to these arrays when giving feedback.

3. Initial Evaluation Experiment

3.1 Experiment Overview

We conducted an initial evaluation of the system: we asked five university students (three males and two females) aged between 21 and 22 to use the system and interviewed them after the experiment to set their feedback. The experiment using human participants was approved by the Life Science Committee of our university.

3.2 Impressions on Recognition Accuracy

The participants provided us with feedback on the recognition accuracy of our system. Judging from the comments and logs, the overall accuracy was approximately 50%.

The first reason for the low recognition accuracy is that the detection accuracy of OpenPose was low. In this system, a web camera was placed at an angle of 45 degrees from the front-left to obtain video recordings about the posture such as leaning forward and backward. However, OpenPose cannot detect whether the sitting posture was perpendicular using only an oblique angle. Adding web cameras to give a frontal and side view or using a Kinect as a substitute may improve the recognition accuracy.

The second reason is due to the number of reference posture models. We compared the interviewee's posture using only one reference model per gender and detected deviations through threshold judgment, we considered that the errors in the detection of postural deviations may have been caused by the fact that the reference model did not take individual differences into account. Accuracy can be improved by preparing several reference models and using one that is as close as possible to the user's height and body size for comparison and detection.

3.3 Impression of Feedback Content

Regarding the feedback content, we received negative comments such as “The feedback (type and number of points) is few.” In fact, each person received feedback approximately two to six times. This may be attributable to the detection accuracy, but could also be due to the weighted priority algorithm used to create the feedback array. Because there may be cases where the detection occurs, but it is not reflected in the feedback sequence, the algorithm needs to be improved. In addition, it was indicated that additional detection points may be necessary. One commenter said, " It would be nice if you could point out finger movements, loudness of voice, speaking etc." indicating the need to point out verbal and nonverbal information.

3.4 Usability of the System

We received positive comments on the usability of the system; e.g., “I don't often have the opportunity to watch my own interview videos, so it's good that they give me advice while watching the videos objectively during the feedback,” “If I could, I would like to practice again,” “The interviewers were very realistic. There was a sense of tension as if it was a real interview, and I felt like I was being watched,” and “I think I can do it alone (could be used by those who are reluctant to go to the employment office for interview practice).”

4. Conclusion

This paper introduces a job interview training system that recognizes the nonverbal behaviors of an interviewee, namely, gaze, facial expression, and posture.

Initial evaluation experiments were conducted, and we identified some points that need improvement: the accuracy was measured at approximately 50% and is not sufficient, the interaction with the CG agent was one-sided and deemed inferior to that with humans, and the number of times feedback was given was approximately two to six times per person, which is too little in terms of volume and content. Improvements will be made to the algorithm for setting the priority of feedback content.

In addition, we will attempt to acquire more nonverbal and verbal information from an interviewee, i.e., hand gestures, prosody, other than gaze, facial expressions, and posture. We need to conduct an experiment with larger number of participants with more variety. For evaluation, quantitative analyses are needed other than qualitative and empirical evaluations.

Demo Video

https://www.youtube.com/watch?v=gBg2BZsFMnQ

REFERENCES

[1] Yuko Matsuda, Minoru Nagasaku, Kunijiro Arai : Infuence of job-hunting anxiety on job-hunting: From the viewpoint of coping, The Japanese Journal of Psychology, 2010, Vol. 80, No. 6, pp. 512-519

[2] Mehrabian A,Silent messages: Implicit communication of emotions and attitudes,Wadworth Publishing.Co.,California,1981

[3] Washburn P.V., Hakel M.D. ,Visual cues and verbal content as influences on impressions after simulated employment interviews.Journal of AppliedPsychology, pp.58,137-140,1973

[4] R. D. Arvey ,J. E. Campion, "The employment interview: A summary and review of recent research," Personnel Psychology, vol. 35, no. 2, pp. 281–322, 1982.

[5] Shogo Okada, Yoshihiro Matsugi, Yukiko Nakano, Yuki Hayashi, Hung-Hsuan Huang, Yutaka Takase, Katsumi Nitta, Estimating Communication Skills based on Multimodal Information in Group Discussions, Journal of the Japanese Society for Artificial Intelligence,Vol.31,No.6, A130-E,2016

[6] MIDAS Information Technology Co., Ltd., https://www.inair.co.jp/

[7] ZENKIGEN Co., Ltd., https://harutaka.jp/

[8] Naim, I., Tanveer, M. I., Gildea, D., Hoque, M. E.: Automated prediction and analysis of job interview performance: The role of what you say and how you say it, IEEE FG, 2015

[9] Rao S. B, P., Rasipuram, S., Das, R., Jayagopi, D. B.: Automatic assessment of communication skill innon-conventional interview settings: A comparative study, ICMI, pp. 221–229, 2017

[10] Nanaho Goda, Keitaro Ishihara, Tomoko Kojiri, Job Interview Support System Based on Analysis of Nonverbal Behavior, IEICE Technical Report,Vol.116,No.517,ET2016-98,25-30,2017

[11] T. Barur, T. Ionut, G. Patrick, P. Kaska, A. Elisabeth.: A Job Interview Simulation: Social Cue-Based Interaction with A Virtual Character, IEEE International Conference on Social Computing (SocialCom2013), pp.220–227, 2013

[12] J. Matthew, B. Laura, F. Micael, J. Neil, A. Michael, J.Emily, W. Katherine, O. Dale, Morris, D, B.: Virtual Reality Job Interview Training for Veterans with Posttraumatic Stress Disorder, Journal of Vocational Rehabilitation 42, pp. 271–279, 2015

[13] H. Tanaka, S. Sakti, Graham. N, T. Toda, H. Negoro, H. Iwasawa, S. Nakamura: Automated Social Skills Trainer, IUI '15 Proceedings of the 20th International Conference on Intelligent User Interfaces, pp. 17–27, 2015

[14] Anderson, K., Andre, E., Baur, T., Bernardini, S., ´ Chollet, M., Chryssafidou, E., Damian, I., Ennis, C., Egges, A., Gebhard, P., Jones, H., Ochs, M., Pelachaud, C., Porayska-Pomsta, K., Rizzo, P., Sabouret, N.: The TARDIS framework: Intelligent virtual agents for social coaching in job interviews, ACE, pp. 476–491, 2013

[15] Damian, I., Baur, T., Lugrin, B., Gebhard, P., Mehlmann, G., Andre, E.: Games are better than books: In-situ ´ comparison of an interactive job interview game with conventional training, AIED, pp. 84–94, 2015

[16] Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B., Picard, R. W.: MACH: My automated conversation coach, UBICOMP, pp. 697–706 ,2013

[17] Langer, M., Konig, C. J., Gebhard, P., Andr ¨ e, E.: ´ Dear computer, teach me manners: Testing virtual employment interview training, International Journal of Selection and Assessment, Vol. 24, No. 4, pp. 312–323, 2016

[18] Baur, T., Damian, I., Gebhard, P., Porayska-Pomsta, K., Andre, E.: A job interview simulation: Social cue-based interaction ´ with a virtual character, SocialCom, pp. 220–227, 2013

[19] Callejas, Z., Ravenet, B., Ochs, M., Pelachaud, C.: A computational model of social attitudes for a virtual recruiter, AAMAS, pp. 93–100 ,2014

[20] Gebhard, P., Baur, T., Damian, I., Mehlmann, G., Wagner, J., Andre, E.: Exploring interaction strategies for virtual ´ characters to induce stress in simulated job interviews, AAMAS, pp. 661–668, 2014

[21] A. N. Joinson. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. European Journal of Social Psychology, 31(2), pp.177–192, 2001

[22] T. Bickmore, A. Gruber, R. Picard. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Education and Counseling, 59(1) pp.21–30, 2005.

[23] MicrosoftAzure, https://azure.microsoft.com/ja-jp/services/cognitive-services/face/

[24] tf-pose-estimation, https://github.com/apulis/tf-pose-estimation

[25] Tobii Technology K.K.,https://www.tobiipro.com/ja/