

My Automated Conversation Helper (MACH): Helping People Improve Social Skills

Mohammed (Ehsan) Hoque
MIT Media Lab
70 Amherst Street,
Cambridge, MA 02139
mehoque@media.mit.edu

ABSTRACT

Ever been in a situation where you didn't get that job despite being the deserving candidate? What went wrong? Psychology literature suggests that the most important skill towards making an impression during interviews is your interpersonal/social skills. Is it possible for people to improve their social skills (e.g., vary the voice intonation, and pauses appropriately; use social smiles, when appropriate; and maintain eye contact) through a computerized intervention? In this thesis, I propose to develop an autonomous and *Automated Conversation Helper* (3D virtual character) that can play the role of the interviewer, allowing participants to practice their social skills in context of job interviews. The *Automated Conversation Helper* is being developed with the ability to "see" (facial expression processing), "hear" (speech recognition and prosody analysis) and "respond" (speech and behavior synthesis) in real-time and provide live feedback on participant's non-verbal behavior.

Categories and Subject Descriptors

J.4 [Social and Behavioral Sciences]: Health

General Terms

Algorithms, Measurement, Performance, Design, Experimentation, Human Factors, Standardization.

Keywords

Training companion, speech processing, facial expressions processing, prosody analysis, feedback.

1. INTRODUCTION

Let us consider John, a junior undergraduate student at MIT - technically gifted and would be a great fit for any technical company. However, despite being the deserving candidate, John has not had any luck in getting an internship of his preference during his junior year. John is disappointed; however, he is also determined to improve his interview skills so that he has better luck for his senior year. John seeks help at the MIT career services and sets up a mock interview with a counselor. After a 15-minute session, the career counselor gives John specific recommendations. These may include maintaining eye contact with the interviewer when necessary. Starting and ending the interview with a social smile, to appear friendly. Varying voice intonation and loudness to express enthusiasm about the position. John returns to his dorm with an understanding of the socially accepted behaviors that he should exhibit during interviews. Yet, he wishes it was possible for him to practice the interview scenarios a few more times and obtain feedback, as he continues to reflect on the advice that he received from the career services. However, John is unwilling to ask his peers for feedback because they are extremely busy, and to some extent, he fears social stigma. Is it possible to help John and many

others like him to improve their interview skills by using an *Automated Conversation Helper*?

In this thesis, I propose to develop an *Automated Conversation Helper*, a new technology that is best described as a virtual agent - an animated and embodied computer generated character that is capable of interacting with its human users (shown in Figure 1). The *Automated Conversation Helper* is being developed with the ability to not only automatically recognize its user's nonverbal behaviors (e.g., facial expressions, body movements, spoken words, and paralinguistic cues), but also synthesize behaviors within a defined scenario. With the ability to see, hear and react, the helper will be able to play the role of a conversational partner, carry on a conversation and provide feedback to its user's nonverbal cues in real-time.



Figure 1. Automated Conversation Helpers, John and Mary

2. BACKGROUND RESEARCH

Even though interaction during interviews are often short and limited, a handful of studies [1][2][3] has provided evidence that interviewers often use that short interaction to predict job performance. Given the limited interaction, interviewers focus on applicant's nonverbal cues to get an understanding of the applicant's personality, social skills and knowledge to determine applicant's suitability. Along with nonverbal cues, there are also knowledge and skills, mental capability, traits, social skills, organizational fit, and preference that are considered towards hiring an applicant. Among other things, Huffcut et al. [1], based on an intensive literature review, have identified that personality traits and social skills were the most frequently rated constructs, followed by mental capability and job knowledge and skills. Some of those social skills include eye contact, frequent smiles, and modulation of speech [4]. For example, as described in [4], an interviewee could appear direct, and honest by maintaining eye contact, or be perceived as more sociable and extroverted by smiling more frequently and varying their voice intonation. Exhibiting the appropriate nonverbal behavior helps to establish rapport with the interviewer, which, according to [5], is the heart of the interview process. Is it possible to develop an *Automated Conversation Helper* that can play the role of the interviewer and can establish rapport with the interviewee through simple contingent nonverbal behavior?

3. PROPOSED RESEARCH

Development of autonomous *Automated Conversation Helper* introduces two unique challenges. First, pushing the boundary of facial expression processing, and speech processing to automatically recognize and understand the nonverbal cues in face to face interactions. Secondly, finding the right medium for the *Automated Conversation Helper* to provide feedback back to the user. The conversation helper could directly provide the feedback through its nonverbal channels (e.g., look disinterested if the interview is not going very well). Another possibility is to provide verbal feedback at the end of the interview similar to standard mock interviews. However, feedback sessions during mock interviews are often very interactive and spontaneous where the interviewee and interviewer engage in somewhat open ended natural conversations. Supporting that level of open ended interactions with an autonomous agent still remains an open problem. Therefore, in this thesis, I have proposed to develop a system that can provide feedback to its users using two different visual formats – immediate and with a time line. In the immediate feedback scenario, participants get to view bird’s eye data of a few selective nonverbal behaviors, as shown in Figure 2. In the upper half of the screen, user gets to view the smile track of his/her interaction across different interview sessions. At the bottom half of the screen, user gets to see comparative analysis of features such as average duration of pauses, speaking rate and amount of weak language (e.g., “like”, “basically”, “I guess” etc.) being used in each session. As a second step, users are given the option to also view the time line format of their interaction as shown in Figure 3. In mock interviews, interviewees often get video-taped and then as part of the feedback process, the counselor and the interviewee watch the video together. To simulate such feature, in time line, interviewees watch their own video along with their entire interaction being visualized through embedded emotional overtones. Some of the emotional overtones include prosodic interpretation of the recognized text. A preliminary diagram is demonstrated in Figure 3, where the vertical displacement between words are presented by pitch, the size of the blobs containing the words represent loudness (bigger the size of the blob, the louder the word was), the colors of the blobs symbolizes smiles (green = really smiling, yellow = might be smiling, red = not smiling).

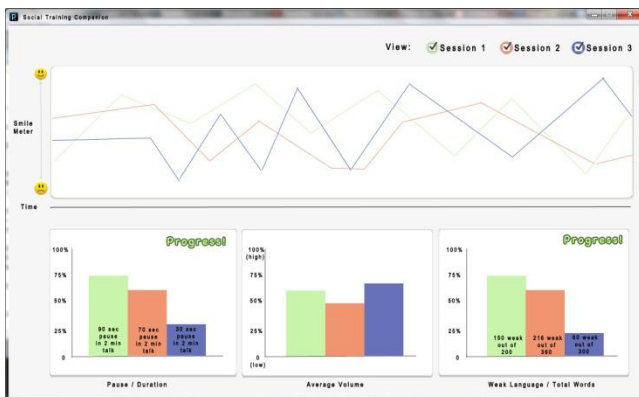


Figure 2: An example of how someone’s nonverbal behavior could be happed to show overall performance

4. HYPOTHESIS

This thesis will primarily investigate whether it is possible to help people improve their interview skills through an *Automated Conversation Helper*. Below are the two hypotheses that will be evaluated in this thesis.

H1: Interaction with an Automated Conversation Helper and then watching his/her own interview video improves interview skills, over self-directed learning of the interview techniques.

H2: Interaction with an Automated Conversation Helper and then watching his/her own video along with a time line view of the nonverbal improves interview skills, over just interaction with the helper and watching one’s own video.

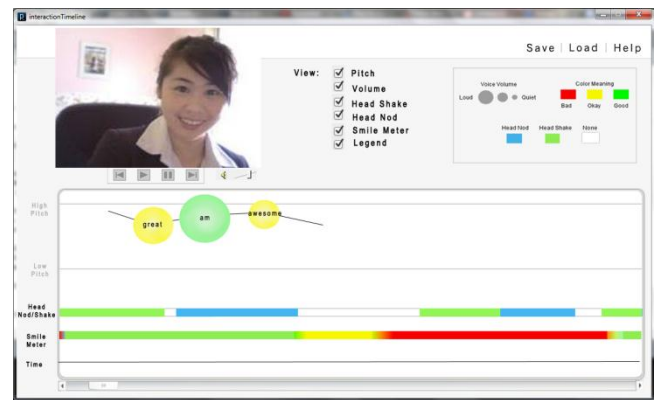


Figure 3. A preliminary example of how an interaction could be automatically transcribed and then visualized according to its non-verbal properties, in real-time.

5. RESEARCH PLAN

The entire study takes place in two phases: study 1 and study 2.

5.1 Study 1

This phase contains job interview scenarios between two individuals – an MIT undergraduate student and an MIT career counselor. Study 1 is more of an exploratory analysis on developing predictive relationships between nonverbal behaviors and job interview skills based on the ratings produced by the counselors as the gold standard. The data generated from Study 1 will be used towards developing models that will drive the interaction between human participants and *Automated Conversation Helper* in study 2. It will also allow for comparison and evaluation of the modeled predictive relationships between nonverbal behaviors and job interview skills with what has been reported in the psychology literature so far. A few possible exploratory research questions that are going to be explored in study 1 are:

- What social behaviors are correlated with interview skills?
- Is it possible to rank social behaviors in order of importance especially in context of job interviews?
- Are there gender effects in interview patterns?

5.1.1 Participants

The interviewers are 4 MIT professional career counselors (3 women and 1 man), who have an average of over 5 years of professional experience as career counselors at MIT and have advanced graduate degrees in professional career counseling.

Interviewees are 28 MIT undergraduates (16 female and 12 male). All participants are MIT undergraduate students in their junior year, and native speakers of English.

5.1.2 Experimental Setup

The experiment was conducted in a room equipped with a desk, two chairs, and two cameras mounted on the wall (Figure 4). Camera 2 was used to record the interviewer and camera 1 to record the interviewee, as shown in Figure 4.



Figure 4: Experimental Setup between the interviewer and the interviewee. Camera 1 is recording the facial expressions and audio of interviewee, whereas Camera 2 is recording the facial expressions and audio of interviewer.



Figure 5: Experiment control room where the experimenter controls the camera to zoom into the participants. The interviewee will be oblivious of the existence of this room.

5.1.3 Procedures

MIT junior students were recruited through flyers and emails in this study. They were told that they would have the opportunity to practice mock interviews with a professional career counselor at MIT and they would get \$10 for just participating. In addition, they were also informed that as part of regular mock interview practice, their interview would be recorded. Each male participant was paired with the male counselor and all the female participants were paired up with the female career counselors. This was done to control for possible gender effects. After the interview, they were informed of the objective of the study and were given the choice to sign the consent form to share their data for research.

5.1.4 Interaction

We spoke to the 4 MIT career counselors and asked them to provide us with 15 questions that are more likely to be asked in any job interview situations, regardless of the position. The 5 questions generated below were the common ones in their list. Below is the exact sequence of the interaction between the MIT career counselor and the participants during study 1.

Counselor: So how are you doing today?(or any question of that sort to initiate the interview)

Participant answers.

Q1. Counselor: (Generic acknowledgement of the answer provided) So, please tell me about yourself?

Participant answers.

Q2. (Generic acknowledgement of the answer provided) Tell me about a time when you demonstrated leadership.

Participant answers.

Q3. (Generic acknowledgement of the answer provided) Tell me about a time when you were working on a team and faced with a challenge, how did you solve that problem?

Participant answers.

Q4. Counselor: (Generic acknowledgement of the answer provided) What is your weakness and how do you plan to overcome it?

Participant answers.

Q5. Counselor: (Generic acknowledgement of the answer provided) Now why do you think we should hire you?

Participant answers.

5.1.5 Debriefing

After the interview ends, the counselor spends a couple of minutes providing feedback to the participant. Then, the counselor walks the participant to the debrief room to the experimenter. The experimenter asks the participant to fill out a questionnaire. In the questionnaire, the participant self-evaluates (on a scale of 1-7) on how well s/he did maintaining eye-contact with the interviewer, how relevant was his/her answers, and how engaging was his/her voice tone, as well as how well s/he did overall in the interview. The interviewer, on the other hand, fills out a similar questionnaire rating the interviewee on speech content, vocal characteristics, and social smiles as well as their individual answers on those 5 questions on a scale of 1 to 7.

5.1.6 Data coding

In study 1, the audio and video files of the participants (i.e., interviewer and interviewee) are being coded. In particular, the facial expressions (different kinds of smiles), vocal characteristics (pitch range, speaking rate, pauses), speech content (structure, qualifications, directness, relevance), speaker presence (enthusiasm, qualifications, directness, relevance), body postures (immediacy cues – orientation, leaning forward, nodding, rocking, fidgeting, mirroring/following gaze, power poses) are now being coded. The coded data will be utilized to develop predictive relationships between nonverbal behaviors and job interview skills.

5.2 Study 2

The purpose of the study 2 will be to evaluate whether social skills can be improved through interactions with Automated Conversation Helper (i.e., validating the hypotheses from Section 4).

5.2.1 Participants

90 Participants, who were not part of study 1, will be recruited for study 2. 90 participants will be randomly divided into 3 groups. There are three phases to study 2: a) Baseline evaluation; b) Intervention; c) Post-intervention evaluation.

5.2.2 Baseline evaluation

The 4 counselors from study 1 will also participate in study 2 and will play the role of interviewers. The interaction between the interviewer and the interviewee and the questionnaire that they would have to fill out will be identical to study 1.

5.2.3 Intervention

Participants will be randomly assigned into 3 groups with equal gender representation. Group 1, consisting of 30 participants, are the control group. After the base line interaction, they are given a standard handout on general tips on how to become successful in interviews. The handout is identical to what MIT students get from the career services after their mock interviews. Group 2 and 3 will consist of 30 participants in each group, who will be part of the conversation helper intervention. As part of their intervention, they will come to the lab and interact with the *Automated Conversation Helper* in an empty room. Participants from Group 2 will get to interact with the Automated Conversation Helper who will engage with the participants and at the end will only show the video of the participant as part of the feedback. With participants from Group 3, the *Automated Conversation Helper* will engage with the participants and at the end will display the video of the interaction as well as the nonverbal behavior as part of a time line.

5.2.4 Post-Intervention evaluation

The interaction in post-evaluation section will be identical to the baseline evaluation. The interviewees will get paired up with the same counselor for mock interviews. The counselor, however, remains oblivious of the fact of whether participant has gone through which part of the intervention. Interviewers will ask the same set of questions. But the questions will be framed differently. It is done to see whether the knowledge learned would generalize when asked the same questions differently.

6. DATA ANALYSIS

Expert assessment: During the baseline, and post-intervention evaluations, we will have measures on how well the participants did according to the ratings of the counselor. The counselor will rate participant's overall performance on eye contacts, engagement in voice tone, relevancy in answers, and being able to smile when necessary. Additionally, the counselor is also going to rate the quality of answer provided for each question.

Self-assessment: During the baseline and post-intervention, the participants are going to rate themselves on the nonverbal cues such as eye-contact, voice tone, relevancy in answers, and being able to smile when necessary. They are also going to rate themselves on their overall performance during the interview.

Third Party assessment: The answers from baseline and post-intervention assessment will be evaluated in random order by others using amazon's Mechanical Turk service. Based on the analysis a set of behavioral variables will be identified for automated analysis. Those behavioral variables, which are indicative of performance in job interviews, will be analyzed and compared before, during and at the end of the intervention.

Automated Behavioral assessment: Based on the findings from study 1 on the predictive relationships between nonverbal Anova analysis will be performed through objective, subjective and behavioral data to investigate changes in performance through the intervention. Other statistical framework, e.g., PARADigm for Dialogue System Evaluation (PARADISE) [6] will also be explored.

7. CONCLUSION

Findings from this thesis could potentially unravel the possibility of using Automated Conversational Helper to improve people's public speaking skills, assist people diagnosed with Asperger's syndrome to improve their social skills, or even automate some aspects of psychotherapy.

8. ACKNOWLEDGEMENT

The author would like to acknowledge Rosalind Picard, Jeffrey Cohn, Bilge Mutlu, and Louis-Philippe Morency, Amy Cuddy and Maarten Bos for providing feedback towards the design of this study. The conversation helper in this thesis is being developed on the existing Multimodal Affective Reactive Characters (MARC) platform developed by Matthieu Courgeon, and Jean-Claude Martin of LIMSI-CNRS. Angela Zhang, Nicole Pan and Anji Ren helped with the data collection.

9. REFERENCES

- [1] A. I. Huffcutt, J. M. Conway, P. L. Roth, and N. J. Stone, "Identification and meta-analytic assessment of psychological constructs measured in employment interviews," *Journal of Applied Psychology*, vol. 86, no. 5, pp. 897-913, 2001.
- [2] M. A. McDaniel, D. L. Whetzel, F. L. Schmidt, and S. D. Maurer, "The validity of employment interviews: A comprehensive review and meta-analysis," *Journal of Applied Psychology*, vol. 79, no. 4, pp. 599-616, 1994.
- [3] P. M. Wright, P. A. Lichtenfels, and E. D. Pursell, "The structured interview: Additional studies and a metaanalysis," *Journal of Occupational Psychology*, vol. 62, pp. 191-199, 1989.
- [4] T. DeGroot and J. Gooty, "Can Nonverbal Cues be Used to Make Meaningful Personality Attributions in Employment Interviews?," *Journal of Business and Psychology*, vol. 24, no. 2, pp. 179-192, 2009.
- [5] J. W. Cogger, "Are you a skilled interviewer?," *The Personnel journal*, vol. 61, no. 11, pp. 840-843, 1982.
- [6] M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella, "PARADISE: A Framework for Evaluating Spoken Dialogue Agents," *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, p. 10, 1997.