Utilizing Real-time Human-Assisted Virtual Humans to Increase Real-world Interaction Empathy

Michael Borish[1], Andrew Cordar[2], Adriana Foster[3], Thomas Kim[4], James Murphy[5], Neelam Chaudhary[6], Benjamin Lok[7]

[1] *University of Florida, USA, mborish@ufl.edu*

[2] *University of Florida, USA, acordar@cise.ufl.edu*

[3] *Georgia Regents University, USA, afoster@gru.edu*

[4] *Georgia Regents University, USA, tkim@gru.edu*

[5] *Georgia Regents University, USA, jamurphy@gru.edu*

[6] *Georgia Regents University, USA, nchaudhary@gru.edu*

[7] *University of Florida, USA, lok@cise.ufl.edu*

Abstract: Empathy is an important aspect of interpersonal communication skills. These skills are emphasized in medical education. The standard source of training is interviews with standardized patients. Standardized patients are trained actors who evaluate students on the effectiveness of their interviews and diagnosis. One source of additional training is interviews with virtual humans. Virtual humans can be used in conjunction with standardized patients to help train medical students with empathy. In this case, empathy training took place as part of a virtual human interaction that represented a patient suffering from depression. However, computers cannot accurately rate empathy, and we thus propose a hybrid experience. We propose a hybrid virtual human approach where hidden workers assist the virtual human. Hidden workers provide real-time empathetic feedback that is shown to the students after their interaction with the virtual human. The students then interview a standardized patient. All empathetic feedback and ratings are based on the Empathic Communication and Coding System (ECCS) as developed for medical student interviews. Fifty-two students took part in the study. The results suggest that students who received feedback after their virtual patient interview did provide more empathetic statements, were more likely to develop good rapport, and did act more warm and caring as compared to the control group that did not receive feedback.

Keywords: Virtual, human, empathy.

1. INTRODUCTION

Empathy is an integral part of a medical student's interviewing skillset. The display of empathy on the part of the interviewer can lead to more satisfying and honest interactions. These improved interactions in turn lead to better medical resolutions (Hojat et al., 2010).

Empathy is the ability to view the world from another person's perspective while identifying their

feelings and experiences. We investigated the impact of hidden workers on the empathy of medical students as they interviewed standardized patients after having interviewed a virtual patient. Empathetic feedback from hidden workers was found to increase several aspects of the standardized patient interaction including empathy, rapport, and caring.

Empathy can be divided into four different forms that are of concern to the medical student or doctor when interviewing a patient: emotive, moral, cognitive, and behavioral. Emotive empathy is the imagining of the patient's emotions. Moral empathy is empathizing with those patient emotions. Cognitive empathy is identification of those emotions, and behavior empathy is conveying understanding of those emotions back to the patient (Teherani, Hauer, & O'Sullivan, 2008).

Empathy is an integral part of a medical student's interview with a standardized patient. A standardized patient is a trained actor who plays the role of a patient with a medical problem and provides constructive feedback.

While useful, standardized patients suffer from drawbacks concerning cost, scheduling, and variability (Stevens et al., 2006) which are not an issue with virtual humans. Virtual humans are playing an increasingly important role in the training of medical students in interpersonal communication skills including empathy. Virtual humans provide a secure sandbox for students to assess patients and receive feedback.

While strides forward have been made in automatic detection of empathy (Mcquiggan, Robison, Phillips, & Lester, 2008), these facets of empathy recognition and feedback are currently best accomplished by humans. Humans can fill in this gap in the underlying algorithms that govern virtual human behavior and can do so without the drawback of traditional standardized patients.

In this paper, we focus on empathetic feedback and the use of a real-time hybrid framework that utilizes hidden workers to assist a virtual human. In our proposed framework, these workers provided ratings for different moments of empathy encountered by the user.

In order to keep the hidden workers busy, they also assisted the virtual human matching algorithm which will be mentioned later for completeness. However, in this paper, we focus primarily on the role of hidden workers' evaluation of empathy, and its resulting impact on learners' empathy with humans.

Using this framework, we conducted a study using hidden workers with a virtual human designed to represent a patient suffering from depression. The workers were trained on both the Empathic Communication and Coding System (ECCS) rating scale (Bylund & Makoul, 2006) and the hidden worker system and were familiar with the scenario.

2. RELATED WORK

Empathy is an important aspect of the doctor-patient relationship. Increased empathy has been shown to improve interpersonal communication and promote therapeutic alliance (Charon, 2001). Increased empathy also results in higher patient satisfaction with their physician and promotes an honest interaction that leads to better medical outcomes (Hojat et al., 2010).

Empathy is regarded as a necessary component to medical student training. However, Hojat et al. (2009) conducted a longitudinal study which showed that as medical student training progresses, the empathy displayed by students decreases.

As part of their education, medical students are exposed to various new technologies. Mohr, Moreno-Walton, Mills, Brunett, and Promes (2011) finds that new medical students are comfortable with the use of new technologies for training. One such technology is virtual humans.

Virtual humans are an additional component to medical student training that can be used for interpersonal communications training. Deladisima et al. (2007) found that virtual humans can elicit the same empathetic response that standardized patients can. Overall, the empathetic response is less with a virtual human than for a real human. However, measured empathy increases the same way it would for a real human interaction. In this case though, the virtual humans did not provide explicit feedback to the students on their empathy performance.

Empathy recognition and its associated feedback are still the domain of humans despite recent improvements (Mcquiggan et al., 2008). So, crowdsourcing techniques can be consulted to allow real humans to be incorporated into the virtual human system.

Humans have previously been utilized to assist computer algorithms in a number of domains. Bigham et al. (2009) used hidden workers to assist in identifying images for the blind using a mobile phone app called VizWiz. Snow, O'Conner, Jurafsky, and Ng (2008) evaluated workers on performing linguistic annotations and found small groups to be on par with expert evaluation. Su, Deng, and Fei-fei (2012) also utilized workers in order to assist in image annotation. All of these various applications illustrate the benefit of human assistance to algorithms. This benefit also applies to the empathy domain in which algorithms struggle to identify what humans can readily rate.

Another similar example is the use of a real-time crowd by Lasecki et al. (2012) in Chorus. Here, humans were used in small groups to assist a Siri-like interface with answering questions. The workers for Chorus were recruited from Amazon's Mechanical Turk marketplace. This is a marketplace for microtasks where workers are paid a few cents to a few dollars to complete some small task.

While recruitment for microtasks from Mechanical Turk was considered, Mason, Street, and Watts (2009) found that payment and duration of task do not indicate a worker's level of skill. Further, the workers for rating empathy must be trained and familiar with a specific coding system. So, local, expert workers were recruited to assist our virtual humans.

3. CONVERSATION SYSTEM

Virtual People Factory (VPF) is a web-based tool for the creation and improvement of virtual humans (Rossen, 2012). VPF facilitates online, text-based virtual human interactions. The VPF interface is shown in Figure 1.

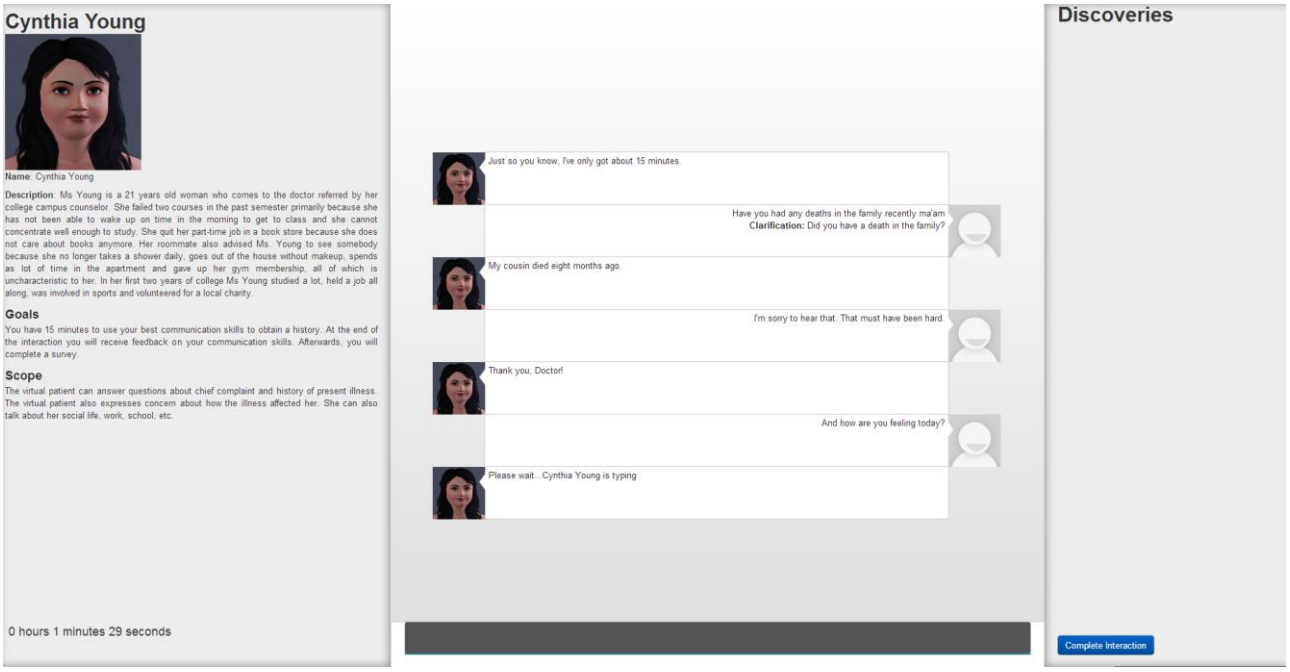

Figure 1: Screenshot of the Virtual People Factory Interface. Using VPF, students conducted a doctor-patient interview using a typed, chat-based interaction.

Here, a user is presented with basic information about the virtual human on the left. On the top-left is a picture of the character, and below that is basic information including the character name and a short description of the case. There is also basic information as to the goals and scope of the interaction. In the center is the main interaction area that contains the transcript and the input to allow a user to type in a question.

Once a user types a question and presses enter, the matching algorithm consults a corpus. A corpus is a structured set of questions and answers that the virtual human can compare its input to and select a response. If the confidence of the response is above a minimum threshold, the related response is returned. Otherwise, the system indicates that it does not have a response for the given input.

There are numerous matching algorithms. Two general classifications of matchers are keyword and Natural Language Processing (NLP) matchers. We have implemented both types of matchers for VPF. We found the accuracy for a keyword matcher is comparable to NLP matchers for virtual humans with a robust corpus and a limited interaction scope such as a medical interview. For this paper, the keyword matching algorithm was used.

A keyword based matcher divides all words into either keywords or stopwords. Keywords are the important words of a sentence that provide meaning such as a name or place and have a relatively lower frequency of occurrence than stopwords. Stopwords are common words that occur often, but don't impart any meaning to sentence such as articles and prepositions.

To find an appropriate response, VPF scans for common words between two phrases and applies appropriate weights to those words. Keywords are weighted more heavily than stopwords. The highest weighted phrase above a minimum threshold is considered a match.

3.1. Hybrid Framework Empathy Alterations

The hidden workers were tasked with providing assistance to two different aspects of the virtual human algorithm. These two aspects were empathetic ratings and matching algorithm assistance. Matching algorithm assistance is outside the scope of this paper as we will primarily focus on the empathetic rating provided by the hidden workers.

The hidden workers first and primary responsibility was to provide ratings for empathetic opportunities during the interaction. An empathetic opportunity is an instance during the interaction when a user has a chance to display empathy towards the virtual human. This opportunity represents the best chance opportunity. This best chance opportunity is the point in time at which a student should absolutely display empathy such as when the virtual human discusses the loss of a family member.

Empathetic opportunities are identified in the system using meta-data tags. The tags are associated with certain responses from the virtual human where users would be expected to offer empathetic words. The empathetic opportunity is thus an opportunistic response from the virtual human and the user input immediately following this response. For example, in the depression scenario we utilized, a student might ask the virtual human about her sister. The virtual human would respond that her sister was recently killed in a car crash. The student would then say how sorry she is to hear that. The empathetic opportunity is the response from the virtual human (the passing of the sister) and the student's input (the extension of empathy by saying she is sorry to hear of the event).

As opportunities are encountered, the relevant response and subsequent input are sent to a hidden worker to be rated. Empathetic opportunities are rated on a scale from 0-6 according to the Empathic Communication and Coding System (ECCS) as developed for medical student interviews (Bylund & Makoul, 2006). The relevant scale and examples are shown in Table 1:

Table 1: ECCS Scale

Level	Description	Example
0 – Denial of Patient Perspective	This response is characterized by the physician ignoring the patient's empathic opportunity or making a disconfirming statement.	Patient: "I'm really concerned about going off this medication." Doctor: "You can go ahead and change into the gown now."
1 – Perfunctory Recognition	This level is characterized by a physician's automatic, scripted-type response to a patient's statement.	Patient: "I was so frustrated." Doctor: "hmmm."
2 – Implicit Recognition	This level contains responses that focus on a peripheral aspect of the empathic opportunity, such as the biomedical issue, but not dealing directly with the progress, challenge or emotion.	Patient: "I'm really happy that I finally quit smoking!" Doctor: "When did you start smoking?"
3 - Acknowledgement	Physician's acknowledgment of the central issue in the empathic opportunity (i.e., conveys that the physician "heard" the patient). Often the	Patient: "I'm so worried about this upcoming biopsy."

	response is a restatement of what the patient has said with or without questions, statements, advice, or offers of help.	Doctor: “I can see that you are scared by the idea of having cancer.”
4 – Acknowledgement with pursuit	Identical to Level 3. The distinction between levels 3 and 4 is whether or not the physician pursues the central issue in the empathic opportunity by asking the patient a question.	Patient: “I’m so worried about this upcoming biopsy.” Doctor: “I can see that you are scared by the idea of having cancer. Tell me more about that.”
5 – Confirmation	The doctors conveys to the patient that the expressed emotional feeling, progress or challenge is legitimate and provides confirmation by using a congratulatory remark, an acknowledgment that the challenge the person is experiencing is difficult, or identifying the importance of the feeling for the patient.	Patient: “I’m so worried about this upcoming biopsy.” Doctor: “Of course you’re scared. Others are scared also, when there’s a possibility of cancer.”
6 – Shared feeling or experience	Explicit statement that he or she either shares the patient’s emotion or has had a similar experience, challenge, or progress.	Doctor: "In my experience, it is hard to study if you have a headache."

In terms of empathetic ratings, medical students traditionally perform poorly. A previous pilot study found that approximately 48% of students’ empathetic opportunities can be classified as a zero with another 28% classified as two. Generally, the higher the rating the better, as higher ratings demonstrate increasing amounts of empathy.

Once users had finished the interaction, all ratings would be displayed to the user for review. An example is shown in Figure 2.

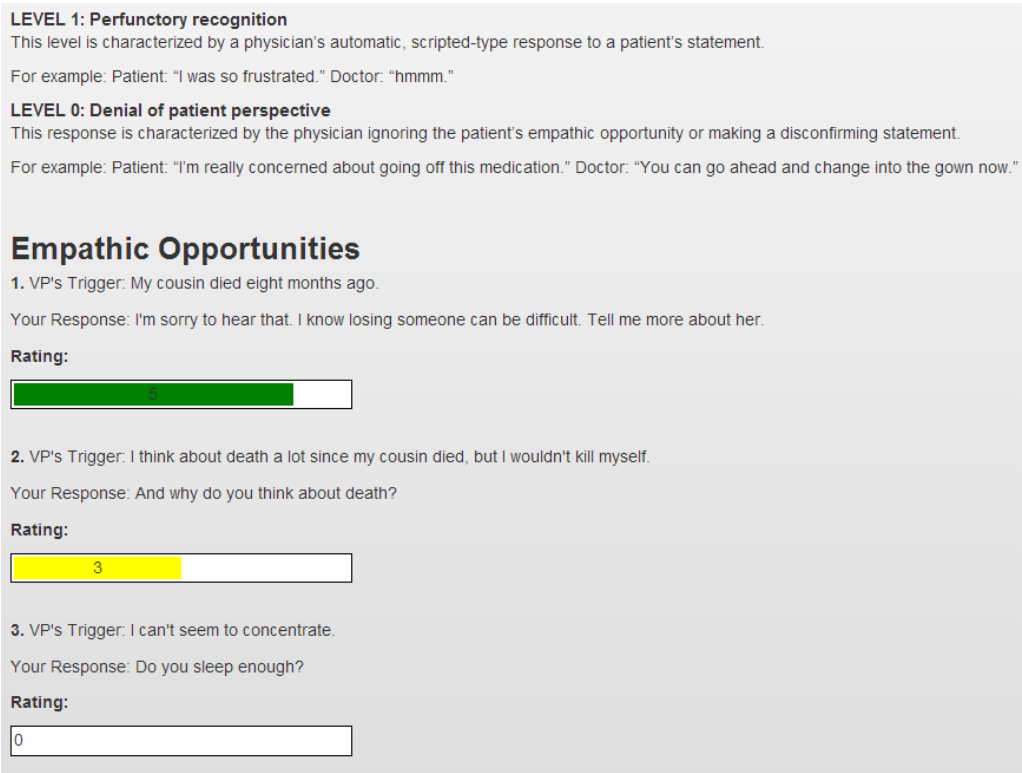

Figure 2: Empathy Feedback Page

Not all rating levels are shown in Figure 2, but the Empathy Feedback page displays a short explanation and description for each rating level at the top of the page. These explanations and descriptions are the same as those listed in Table 1. The webpage also lists every empathetic opportunity and its rating beneath. This way a user could review how empathetic they were during their interaction with the virtual human.

As part of this framework, it is critical to keep background workers occupied. Background workers need incentives to stay engaged so as to be available during empathetic moments. To keep background workers engaged, we leveraged workers to assist the keyword matcher in selecting paraphrases when the matching algorithm was not able to clearly match a student's input to the corpus. The closest paraphrases to a question are called speech clarifications. Speech clarifications occur when no paraphrase scores above a certain threshold calculated by the matcher. The speech clarifications are the best choices for an alternative question because they are the highest scored alternatives in the conversational corpus. In essence, background workers assist the virtual human by helping the matching algorithm throughout the interaction. Workers assist the matching algorithm approximately 40% of the time. Besides utilizing this unused time, the clarifications form the basis for future improvements to the virtual human (Rossen, 2010).

3.2. Hybrid Approach Considerations

In order to give the hidden workers the time necessary to complete their work, a small delay was introduced into VPF. This delay hid the latency of the workers as they provided real-time assistance for empathy ratings and clarifications. The delay was present for all responses the user received from the virtual human.

For instances in which no hidden worker help was required, the delay is based on the length of the response and the average typing speed of the typical user. Otherwise, the delay is the amount of time taken by hidden workers to provide a response or a time limit of seven seconds is reached. This time limit is based on a preliminary study that found five seconds of delay does not impact overall virtual human perception. Given the findings from this study, the time limit of the delay was increased slightly to seven seconds.

By always having a response delay, the system maintains the illusion of a chat with another person. Framing the interaction as an online chat, as well as hiding the delay, is accomplished with displaying a "Character name is typing…" message while the user is waiting for a response. A typical example is shown in the bottom virtual human text box in Figure 1.

3.3. Worker Interface

Input from the user on the interaction page streams to the workers in real-time to a separate webpage. The sections of the webpage that workers use to provide hidden assistance to the virtual human is shown below.

The upper section of the worker page allows a worker to select a clarification as shown in Figure 3.

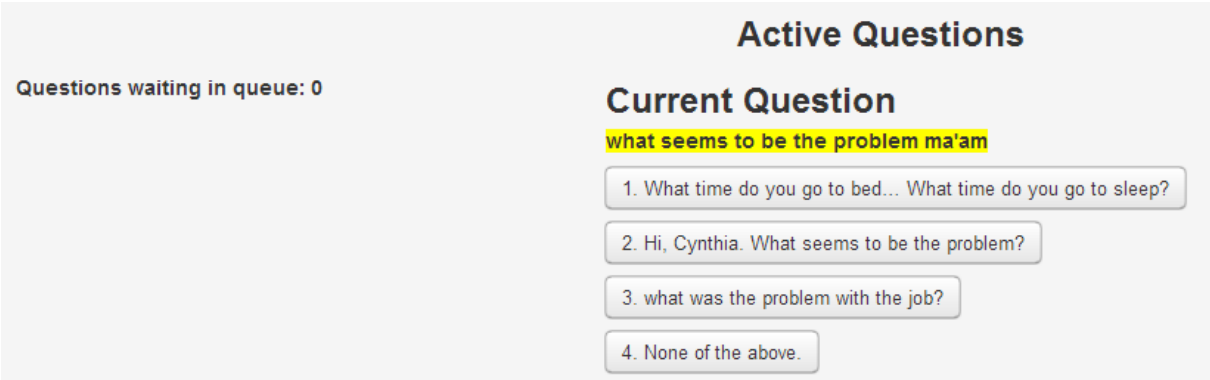

Figure 3: Clarification Selection Area

The lower section of the worker page is shown in Figure 4.

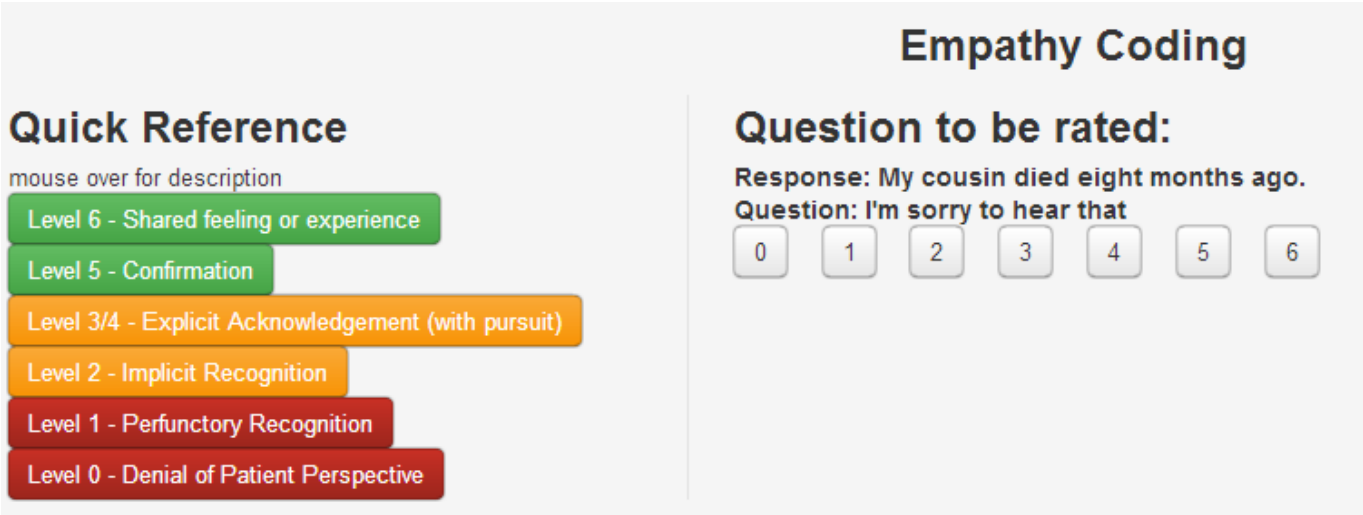

Figure 4: Empathy Selection Area

This section allows the worker to provide the empathetic ratings that are shown to the user on the empathetic feedback page as previously described in the implementation section. Once a user receives a response that is tagged as an empathetic opportunity, the system sends across this response as well as the next user input.

This empathetic opportunity is then rated according to the ECCS scale. Each level is listed on the left and provides a short description for the worker to reference if necessary. This description is based on the scale's full description listed in Table 1.

A typical example of student response for an empathetic opportunity is shown in Figure 4. Here, the virtual human has just responded that her cousin died eight months ago. The student's next input is that they are sorry to hear that. Based on the ECCS scale, this pair of sentences qualifies as Level 3. The student explicitly acknowledged the virtual human's emotional response, but did not continue with a follow up question to probe for more information.

4. STUDY DESIGN

We conducted a study to investigate the impact of the proposed framework and the use of hidden workers to provide empathetic feedback. All students were 1st year medical students that had not previously been exposed to our system nor had they taken part in a previous study. The medical students (N=52) were randomly divided into two groups. The study flow is shown below:

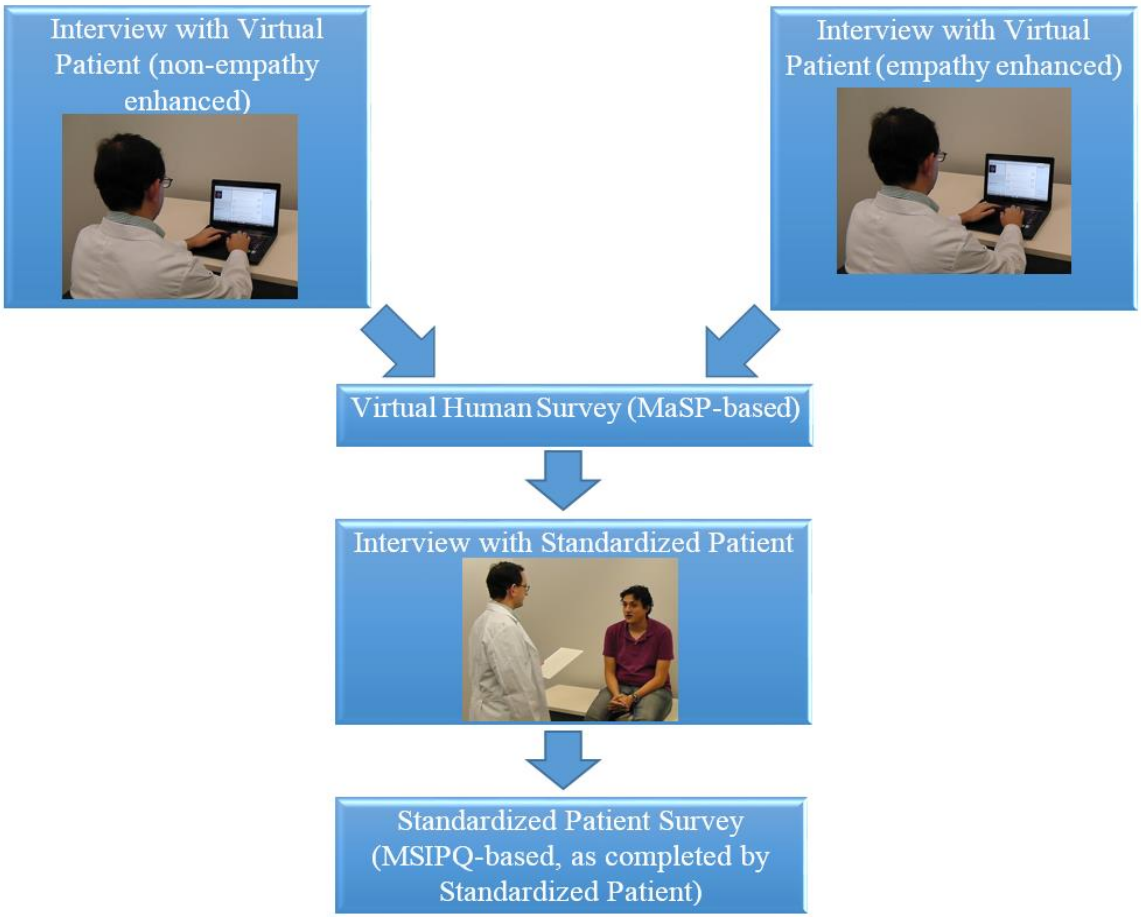

Figure 5: Study Flow

The first group is the control group (N=17). The control group engaged our depression virtual human in an interaction without the aid of a hidden worker or empathetic feedback. Once their interaction was complete, they filled out a survey concerning the virtual patient. Afterwards, the students engaged a standardized patient who also represented a depression scenario.

The second group is the intervention group (N=35). The intervention group engaged our depression virtual human with the benefit of a hidden worker. They then completed the same survey as the control group and following this, received empathetic feedback as described in the implementation section. Afterwards, the students also engaged a standardized patient representing a depression scenario in the same way as the control group.

Both groups filled out a survey after their interaction with the virtual human. This survey is based on the Maastricht Assessment of Simulated Patients (MaSP) (Wind et al., 2004) survey for virtual human evaluation. All responses were on a 5-point scale of Poor, Fair, Average, Good, and Excellent and consisted of the following questions:

- The virtual patient responded in an appropriate amount of time.
- Did the virtual patient interaction help you learn how to formulate questions about a patient's history?
- Were the virtual patient's answers appropriate?
- Did the interaction with the virtual patient simulate real life?
- The virtual patient responded to questions in a natural manner.
- The virtual patient appeared to withhold information.
- The virtual patient communicated how he/she felt during the session.
- The virtual patient stimulated me to ask questions.
- The virtual patient understood my questions.
- Did you enjoy this interaction?
- How valuable is this interview training tool?
- Was Virtual Patient Factory easy to use?
- Were the discoveries useful?
- Was the transcript useful?
- How do you rate the overall interaction?

Both groups also had their overall empathetic rating calculated for the virtual human interaction. The empathetic rating for the interaction was calculated based on the cumulative rating for all empathetic opportunities that occurred during the interaction.

The workers that assisted the virtual humans in the intervention group were all experts. They were trained on the appropriate classification of students' empathetic opportunities and achieved an intra-class correlation coefficient (ICC) of .81. ICC is a measure of inter-rater reliability (IRR) which is how well different raters agree on what the classification of a particular empathetic opportunity should be. A minimum of .7 is considered acceptable in various medical disciplines.

The standardized patients that were interviewed for both groups were blind to the study condition. After their interaction with the students, the standardized patients completed a survey, the Medical Student Interviewing Performance Questionnaire (MSIPQ) (Black & Church, 1998). The MSIPQ is a well-established checklist that rates the medical students' professional appearance, behavior, empathy, and rapport. The relevant section of the questionnaire included the following questions:

- The examinee offered encouraging, supportive, and/or empathetic statements.
- The examinee developed a good rapport with me.
- The student appeared warm and caring.
- The examinee knocked first before entering the exam room.
- The examinee addressed me by my name
- The examinee introduced him/herself
- The examinee washed his/her hands
- The examinee began the interview with open-ended questions
- The examinee repeated, and/or summarized the information I gave him/her during the interview.
- The examinee used language appropriate for me as a person (did not used "medical jargon").
- The examinee's overall professional appearance was appropriate.
- The examinee did not display judgmental behavior.
- The examinee conducted the interview in an organized manner.
- The examinee did not invade my personal space.
- The examinee demonstrated attentive listening.
- The student seemed stiff and unnatural.
- The student seemed confused and nervous.
- The student seemed comfortable talking to me.
- The student talked down to me.

The video for the standardized patient interaction was also reviewed by experts. Empathetic ratings were applied for each empathetic opportunity in the same way they were applied by hidden workers during the virtual patient interaction. The overall empathetic rating was calculated from these opportunities.

5. RESULTS AND DISCUSSION

5.1. Standardized Patient

Students demonstrated more empathy in their interview with a standardized patient after having interacted with a virtual human that provided empathy feedback than students who interacted with a virtual human that did not provide feedback.

For the questions presented as part of the MSIPQ to the standardized patients, a Chi-Squared test was performed. Statistical significance was found for the following statements: “The examinee offered encouraging, supportive, and/or empathetic statements”, “The examinee developed a good rapport with me”, and “The student appeared warm and caring.”

In the depression standardized patient communications survey, those in the intervention group were more likely to have offered encouraging, supportive, and/or empathetic statements (df=1, p<0.0001), were more likely to have developed a good rapport (df=1, p=0.003), and were more likely to appear warm and caring (df=4, p=0.0157) than those in the control group. The frequencies for those ratings are shown respectively in Figure 6.

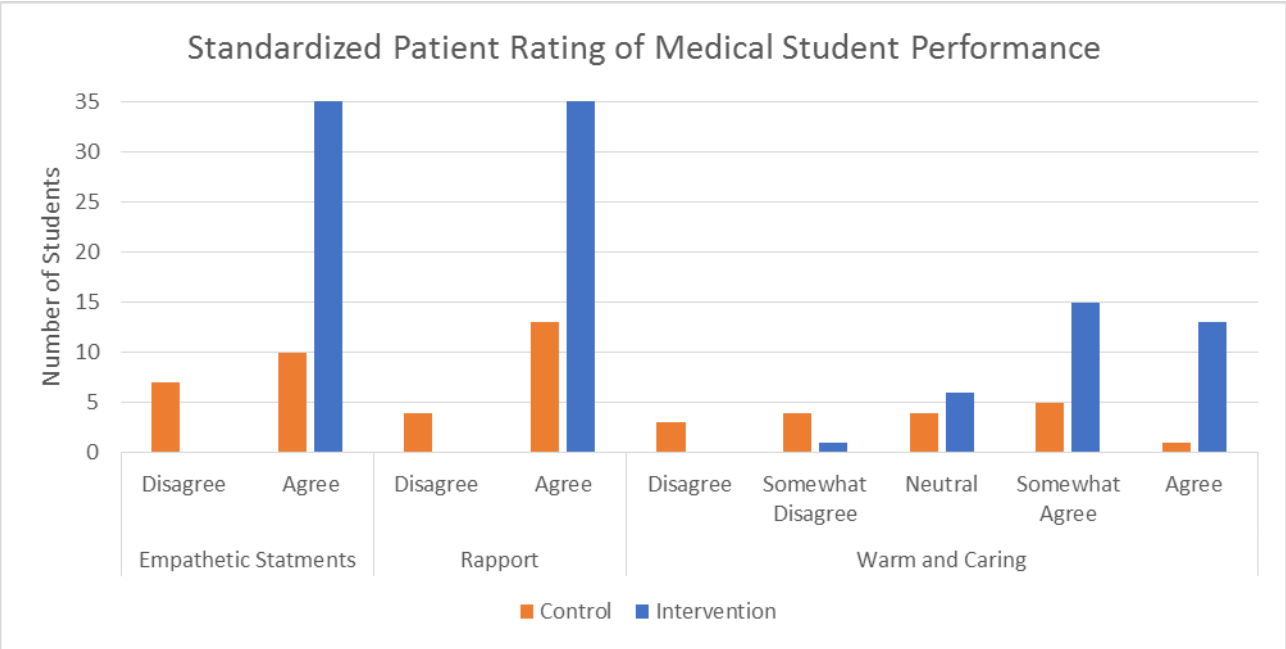

Figure 6: Standardized Patient Ratings for Medical Students

This relationship held true when accounting for the number of empathetic opportunities that students encountered when interviewing the depression virtual patient.

These three specific statements are directly related to how empathetic and caring the medical students appeared to the standardized patients. All three indicate a better interaction as scored by the standardized patient demonstrating that the empathy feedback page after the virtual human interaction influenced the student’s behavior in the mock scenario. The magnitude of this

difference is easily visible in the frequency counts from the standardized patient. For example, for the section on empathetic statements in Figure 6, the standardized patient agreed with the statement that “The examinee offered encouraging, supportive, and/or empathetic statements ”for all 35 of the students in the intervention group. However, only 10 students out of 17 from the control group were listed as such.

The standardized patient interaction video was reviewed by experts and the empathetic opportunities scored on the ECCS scale. Statistical significance ($t=2.39$, $p < .0206$) was found between the control (mean = 2.29, SD = .67) and intervention group (mean = 2.87, SD = .73). This indicated that the medical students were more empathetic after having interviewed a virtual human first.

The data show that the medical students have had their behavior altered by interacting with the virtual human before the interview with the standardized patient. Allowing a student to review how empathetic they were to the virtual patient through the feedback page is enough to not only increase several specific measures of standardized patient interaction efficacy, but the overall amount of empathy as well.

An ANCOVA analysis was conducted to determine the influence of a variety of covariate factors on the empathetic score analysis. The base ANCOVA model controlled for the number of empathetic opportunities encountered during the interaction while covariates for gender, race, medical specialty, and mental health experience score where investigated. No statistical significance was found for any covariate.

This reinforces our conclusion that the increase in empathy was due to the feedback received as part of the virtual human interaction

5.2. Virtual Human

Students demonstrated more empathy when interviewing a virtual human that provided empathy feedback than with a virtual human that provided no empathy feedback.

An ANOVA was performed to determine the statistical significance of overall virtual patient empathy. Statistical significance was also found for the overall virtual patient empathetic ratings ($F=4.98$, $p=.029$). Students in the intervention group (mean = 2.2, SD = .32) had higher empathetic ratings than those in the control group (mean = 1.9, SD = .32).

A difference in the empathy among students was not expected during the virtual patient interaction. At this point, no student had received any empathetic feedback. Due to IRB restrictions and educational goals, as discussed later in the limitations section, the students did have an idea as to the overall nature of this exercise. However, all were informed to the same degree and blind to the condition they were grouped in. As a result, we expected that the students in both groups might be more empathetic in general. Regardless, the intervention group performed better than the control group, and we can only hypothesize as to the reason for a difference between groups.

We hypothesize that this difference could be due to two factors. First, the issue of priming. Students could have been primed to be more empathetic due to IRB limitations and the stated educational goal of improving their empathy. Second, the intervention group had more self-reflection time due to the latency of the hidden worker assistance.

Despite these factors, it should be noted that there is a difference between statistical and meaningful significance. While statistical significance was found, both averages are still around

two. This is a low score on the ECCS scale and both groups could improve their empathy significantly. This significant increase in empathy is what was found in the intervention group as previously described in the standardized patient results. Still, this unexpected difference presents an opportunity for additional further study.

A t-test was conducted on the virtual human response survey. Statistical significance ($t = 3.209$, $p = .002$) was found for the statement "The virtual patient responded in an appropriate amount of time." Students in the control group rated the response time of the virtual human higher (mean = 4.74, SD = 0.62) versus the intervention group (mean = 4.09, SD = 0.98).

This difference in perception is likely due to the increased latency involved in receiving a response. As mentioned previously, when a clarification is sent to the hidden workers, there was a timeout of seven seconds. During this time, the user receives a status message that the character is typing. A previous study used a similar worker page to allow hidden workers to answer clarifications as part of a small group where majority vote was required. This scheme allowed only a five second timeout. That study identified no difference in perception. So, using this technique of displaying a status message to hide the latency has an upper bound between five and seven seconds before negative perception of the virtual human occurs. Despite thinking that the virtual human should respond quicker, no other change in perception was found.

6. LIMITATIONS

While we took as much care as possible in designing the study and controlling for factors that may have influenced results, there were some limitations. Due to student availability, we are unable to use the same students for repeat experiments. Repeat experiments with the same students would be necessary to determine what type of lasting effect, if any, the virtual human interaction has on the students.

Due to IRB restrictions, the students were aware that the study involved empathy. Empathy was mentioned on the recruitment flyer. This likely biased the students and led them to be more empathetic than they otherwise would normally be. However, this same flyer was shown to all participants.

Additionally, participants in the intervention group were reminded to review any feedback the virtual human provided. This too likely biased them to be more empathetic with the standardized patient. However, all students are always aware of the standardized patient checklist and know that empathy is a major component.

7. CONCLUSION AND FUTURE WORK

We found the overall empathetic response from students when interacting with a standardized patient was improved after interacting with a virtual patient that utilized our hybrid framework. This hybrid framework makes use of hidden human workers to assist the virtual human. The hidden workers provide ratings for empathetic opportunities to provide feedback to the students.

This feedback altered students' empathetic response. Empathy is a vital component to medical student training. Improved empathy improves doctor-patient interactions and can make a significant difference in the treatment and outcome for a patient. Despite improvements shown here, the overall empathetic rating continues to be fairly low on the ECCS scale with plenty of room for improvement.

Regardless of the low initial scores, as a result of these interactions, 52 medical students have now received additional training when dealing with a patient suffering from depression. Depression is a serious condition that medical literature indicates is hard to treat due to training constraints and stigmatization. Medical educators are thrilled to have a new tool that results in demonstrable improvements in empathy to help train students on a previously difficult subject area (Stevens et al., 2006).

We intend to pursue this research with additional studies. We intend to see if some of the worker's responsibility can be shifted to an automated algorithm now that there is a basis for classification. Also, the virtual humans have a difficult time in responding to the empathetic opportunities as opposed to general question matching. We intend to investigate whether direct worker involvement at those specific moments would lead to an improved interaction. We also intend to investigate the underlying causes for the difference in initial empathy displayed by the different groups.

ACKNOWLEDGMENTS

The authors would like to thank everyone involved for their work in helping to make this study a success. This work was funded by the Arnold P Gold Foundation research grant FI-12-010.

REFERENCES

Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., … Park, C. (2010). VizWiz : Nearly Real-time Answers to Visual Questions.

Black, A. E., & Church, M. (1998). Assessing medical student effectiveness from the psychiatric patient's perspective: The Medical Student Interviewing Performance Questionnaire. *Medical Education*, *32*(5), 472–478. doi:10.1046/j.1365-2923.1998.00247.x

Bylund C, Makoul G, *Empathic communication and gender in the physician-patient encounter,* Patient Education and Counseling 2002, 48: 207-216

Charon, R. (2001). Narrative Medicine. *JAMA*, *286*(15), 1897–1902.

Deladisma, A. M., Cohen, M., Stevens, A., Wagner, P., Lok, B., Bernard, T., … Lind, D. S. (2007). Do medical students respond empathetically to a virtual patient? *American journal of surgery*, *193*(6), 756–60. doi:10.1016/j.amjsurg.2007.01.021

Hojat, M., Louis, D. Z., Maxwell, K., Markham, F., Wender, R., & Gonnella, J. S. (2010). Patient perceptions of physician empathy, satisfaction with physician, interpersonal trust, and compliance. *International Journal of Medical Education*, *1*, 83–87. doi:10.5116/ijme.4d00.b701

Hojat, M., Vergare, M. J., Maxwell, K., Brainard, G., Herrine, S. K., Isenberg, G. A., … Gonnella, J. S. (2009). The Devil is in the Third Year : A Longitudinal Study of Erosion of Empathy in Medical School, *84*(9), 1182–1191.

Lasecki, W.; Kulkarni, A.; Wesley, R.; Nichols, J.; Hu, C.; Allen, J.; and Bigham, J. 2012b. Chorus: Letting the crowd speak with one voice. In University of Rochester Technical Report, 1–10

Mason, W., Street, W., & Watts, D. J. (2009). Financial Incentives and the “ Performance of Crowds ”, *11*(2), 100–108.

Mcquiggan, S. W., Robison, J. L., Phillips, R., & Lester, J. C. (2008). Modeling Parallel and Reactive Empathy in Virtual Agents : An Inductive Approach, (Aamas), 167–174.

Mohr, N. M., Moreno-Walton, L., Mills, A. M., Brunett, P. H., & Promes, S. B. (2011). Generational influences in academic emergency medicine: teaching and learning, mentoring, and technology (part I). *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, *18*(2), 190–9. doi:10.1111/j.1553-2712.2010.00985.x

Rossen, B., Cendan, J., & Lok, B. (2010). Using Virtual Humans to Bootstrap the Creation of Other Virtual Humans, 392–398.

Rossen, B., & Lok, B. (2012). A crowdsourcing method to develop virtual human conversational agents. International Journal of Human-Computer Studies, 70(4), 301–319. doi:10.1016/j.ijhcs.2011.11.004

Snow, R., O’Conner, B., Jurafsky, D., & Ng, A. Y. (2008). Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (pp. 254–263)

Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raij, A., Harrison, C., … Lind, D. S. (2006). The use of virtual patients to teach medical students history taking and communication skills. *American journal of surgery*, *191*(6), 806–11. doi:10.1016/j.amjsurg.2006.03.002

Su, H., Deng, J., & Fei-fei, L. (2012). Crowdsourcing Annotations for Visual Object Detection.

Teherani, A., Hauer, K. E., & O’Sullivan, P. (2008). Can simulations measure empathy? Considerations on how to assess behavioral empathy via simulations. *Patient education and counseling*, *71*(2), 148–52. doi:10.1016/j.pec.2008.01.003

Wind, Lidewij A., et al. "Assessing simulated patients in an educational setting: the MaSP (Maastricht Assessment of Simulated Patients)." Medical Education 38.1 (2004): 39-44.