High Score! - Motivation Strategies for User Participation in Virtual Human Development

Shivashankar Halan[1], Brent Rossen[1], Juan Cendan[2], and Benjamin Lok[1]

[1] Computer Information Sciences Engineering,
[2] Department of Surgery,
University of Florida, USA
shivashankarh@ufl.edu,{brossen,lok}@cise.ufl.edu,
cendajc@surgery.ufl.edu

Abstract. Conversational modeling requires an extended time commitment, and the difficulty associated with capturing the wide range of conversational stimuli necessitates extended user participation. We propose the use of leaderboards, narratives and deadlines as motivation strategies to persuade user participation in the conversational modeling for virtual humans. We evaluate the applicability of leaderboards, narratives and deadlines through a user study conducted with medical students (n=20) for modeling the conversational corpus of a virtual patient character. Leaderboards, narratives and deadlines were observed to be effective in improving user participation. Incorporating these strategies had the additional effect of making user responses less reflective of real world conversations.

Keywords: Virtual Humans, Intelligent Agent Authoring Tools, motivational strategies, persuasive technology, virtual patients, leaderboards, narratives, deadlines.

1 Introduction

End-users are increasingly becoming an integral part of application development processes. User interactions with the system during the development phase have different goals than interactions after deployment. The developmental user interactions primarily benefit the developers by helping them identify issues with the system. This is opposed to post-deployment interactions where the goal is to use the system to primarily benefit the user. During the developmental interactions, the system will likely have poor performance as measured by number of errors and ease of use. The poor performance of the system causes a *motivation gap.* We define a motivation gap as the lack of proper motivation for a user to participate in a particular task. Since the developmental interactions do not primarily benefit the user, there is minimal motivation for the user to expend effort and time to interact with a system that is still in development. We propose using external motivation strategies to address the motivation gap.

In the past, providing monetary compensation and extra credit have been effective motivational techniques for increasing user participation [3] [4], however, these motivational techniques are not always applicable. For example, providing monetary

J. Allbeck et al. (Eds.): IVA 2010, LNAI 6356, pp. 482–488, 2010.

compensation can be costly and may bias participants. Video games and social networking platforms have been successful in bridging the motivation gap [1], and thus we searched their literature for easily deployable motivation techniques that can effectively persuade users to participate in developmental interactions.

Our work focuses on creating virtual human patients for training healthcare students with their interviewing and interpersonal skills. The development of these virtual patients requires participation from healthcare students in the developmental interactions. We enlisted the help of healthcare experts to select appropriate motivation strategies for motivating participation amongst the healthcare students. When presented with various possible motivation strategies, the healthcare experts identified three strategies that they anticipated would work well for the specific application of motivating medical students: leaderboards, narratives and deadlines.

- A **leaderboard** is an ordered display of the scores and the names of participants in a competition. Video games are a good example of where leaderboards have been used to encourage user participation [1].
- A **narrative** is the telling of a story or account of events or experiences [2].
- A **deadline** is a specific timeframe within which the user is expected to complete a task.

In this paper, we evaluate the applicability of leaderboards, narratives and deadlines in bridging the motivation gap and persuading user participation in the developmental interactions for VH corpus development. A user study was conducted with medical students, and the results indicate that leaderboards, narratives and deadlines were effective in increasing user participation amongst the medical students.

2 Motivation Strategies for Conversational Modeling

One method of simulating human-VH interpersonal conversations is to use a corpus based retrieval approach. The corpus of a VH consists of question-answer pairs of what the users will say to the VH (stimulus) and what the VH will say back (response). When a user asks a question, the system searches the corpus for the most similar question and provides the paired answer. Developing the corpus for human - VH interpersonal conversations is a time-consuming task [3]. This need for extended time commitment is because of the need to gather the various stimuli that can trigger a response. For example, the response "*I have really bad stomach pain*" could be a response to any of the following stimuli: "*How are you feeling?*", "*What's wrong*?" and "*How can I help you?*". The corpus requires comprehensive coverage of the questions that will be asked and the answers the VH should provide.

The time taken for corpus development can be reduced by using a technique called Human-Centered Distributed Conversational Modeling (HDCM) [4]. In HDCM, novice users conduct interactions with the VH during the VH corpus development phase and pose questions (stimuli) to the VH. Stimuli to which the VH did not have a response are automatically gathered into a list. The user is further enlisted to mark when the VH responds incorrectly and these are gathered into the same list. The domain expert can then go through this list to validate new stimuli for existing responses or add new responses for the new stimuli. Thus, through iterations of user interactions

and subsequent expert reviews, the corpus of the VH is progressively developed. The HDCM method was observed to take only 10% of the time taken for conversational modeling by previous approaches [4].

As an example, the corpus of a VH that plays the role of a patient suffering from lower back pain was developed for training medical students. The first draft of the corpus that was developed by a physician contained 51 responses and 139 stimuli. After 27 iterations of different users interacting followed by expert reviews of the accumulated stimuli, the corpus now has 118 responses and 446 stimuli. The HDCM model saves time [4] but shifts much of the corpus development work from the expert to the novice. This leaves domain experts with the problem of motivating enough novice users to interact with the VH during the corpus development phase.

3 Motivation Strategies for Medical Students

The target application for this work is the creation of VHs that act as virtual patients. The corpus for these virtual patients is built by medical educators, and medical students constitute the set of novice users who need to interact with the virtual patient during the corpus development stage. The participation of medical students in previous iterations of developmental interactions has been suboptimal and has been observed to vary from 50% to as low as 5%. This low response rate can be attributed to the fact that the subjects of our research are medical students in their surgical clerkship phase. This is a particularly difficult time in a medical student's professional career marked by long work days, patients with high acuity of illness and replete with high-stakes decision-making. It is in this environment that we now ask them to take on another task – to participate in the developmental interactions for conversational modeling of a virtual patient VH. As a result, it is difficult to get the medical students to participate in the VH developmental interactions. Also, requests for participation sent out during advance stages of the semester, which is a very busy time for the medical students, received fewer responses.

The experts' reasoning for choosing the three motivational strategies discussed includes the desire to leverage the competitive nature of medical students for the leaderboards. Medical education also engages the student by presenting medical knowledge in the context of the patient [5] and thus we utilize the narrative format to continue this tradition. Medical students are readily engaged when the information is couched in a clinical milieu. Medical students are also well trained to react to critical timelines and focus their attention on tasks with imminent deadlines.

4 Implementation

The external motivation strategies evaluated in this paper were implemented using Virtual People Factory (VPF). VPF is a web-based application that allows users to create and interact with VHs [4]; the VHs created using VPF conduct natural language conversations using a corpus retrieval approach.

Leaderboards: During an interview that aims at bettering interpersonal skills, there generally exists a mechanism to evaluate the trainee's performance. In the case of

medical interview training, a checklist of information that is critical to the scenario is used for evaluating the student's performance. Each piece of information on the checklist is called a "discovery" because when the user asks the appropriate question, they "discover" a critical piece of information- (e.g. "*patient smokes 2 packets of cigarettes a day*"). In our implementation, the user's score is based on how many discoveries were elicited during the interaction. This score is then displayed on the leaderboard. A sample leaderboard interaction is shown in Fig.1. Users are allowed to have multiple interactions with the VH, and the score for a user is the highest score over all interactions with the VH (maximum score is 100).

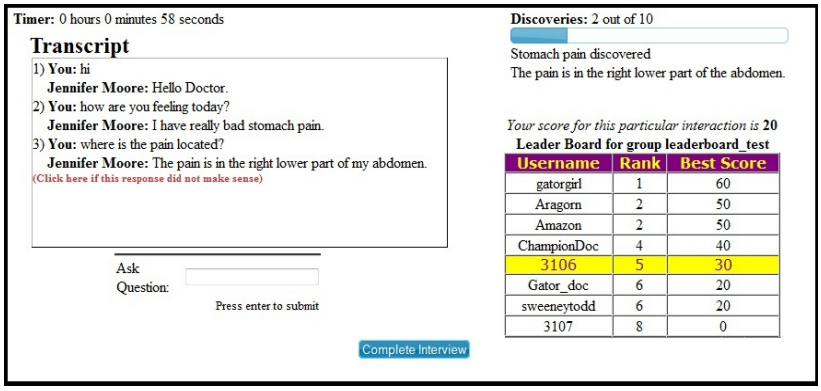

Fig. 1. Leaderboard enabled medical student–virtual human interaction

The users were given the option to choose nicknames for themselves that will be displayed on the leaderboard. For users who did not choose a nickname, a numerical user id was displayed on the leaderboard to represent the user. This provides anonymity for students who would prefer others did not know how they performed.

An important part of appealing to the competitive nature of the participants was to constantly inform them of their leaderboard status. The leaderboards show the user's current position throughout the interaction. Through this mechanism, students receive immediate feedback in seeing their scores pass that of colleagues and classmates. In addition to displaying their status during the interaction, three different kinds of e-mails were sent to the users that informed them of their status on the leaderboard: after an interaction with the VH, weekly update e-mails and whenever a user's score was beaten by another user.

Narratives were a part of the participation-request e-mails that were sent to the users. Medical students who conduct medical interviews are normally given a patient case document that contains information about the demographics and background of the patient. This case document was used to create the narrative for the VH interaction. A sample excerpt from the narrative for a virtual patient suffering from right lower quadrant pain is shown below (used in the evaluation study in Section 5).

> "*Jennifer Moore, a 21-year-old female, presents to the emergency department complaining of lower right quadrant abdominal pain. . The onset of the pain was 1 p.m. The current time is 5 p.m. She has never had this kind of pain before and is extremely worried about what might be the cause.*"

Deadlines add a sense of urgency and appeal to the medical student's instinct for providing immediate medical attention. The deadline for the interaction is included in the invitation e-mail sent to the user. Additionally, update e-mails also reference the remaining time for the user to meet with the VH.

5 Evaluation Study

We present the results of an experimental user study conducted with (n=20) medical students. The study was a between-subjects experimental design with the use of external motivation as the independent factor. The participants were split randomly into two groups:

Control group: The control group with n=9 participants interacted with the VH without the leaderboards, narratives or a deadline.

Experimental group: The experimental group with n=11 participants was sent an e-mail with a narrative and a two-week deadline. This group also saw leaderboards during their interactions and was sent update e-mails three times a week.

The chosen medical scenario was one where the VH was suffering right lower quadrant pain because of an ectopic pregnancy. One of the reasons for choosing this particular medical scenario was that we had difficulties in the past recruiting users for participating in the developmental interactions for this very same scenario. At the time of testing, the VH's corpus consisted of 115 responses and 745 different stimuli for these responses.

Hypothesis-1: Participants who are sent invitations to interactions with leaderboards, narratives and deadlines will be more likely to participate in the developmental VH interactions than participants without leaderboards, narratives and deadlines as measured by the response percentage of the participant group to recruitment e-mails.

Hypothesis-2: Participants who are sent invitations to interactions with leaderboards, narratives and deadlines will spend a greater amount of time interacting with the developmental VH than participants without leaderboards, narratives and deadlines.

The following metrics were evaluated:

- **Number of interactions** – the total number of VH interactions. Students performing multiple interactions were counted multiple times.
- **Duration of the interactions** - the time from the start of the interaction to when the users clicked "complete interview"

5.1 Results

The control group received one interaction out of nine users (11%). The experimental group had fifteen interactions out of eleven users. Seven of the eleven users

conducted one interview, while one user conducted three interviews, and another conducted five for a total of nine users responding out of eleven (**82%**).

Since the control group received only one interaction, we examined data from a previous developmental study group (n=22) that used the same scenario as the current study (*previous group)*. Of the 22 participants recruited for the previous group, 12 were contacted in person by their professor and requested to participate but still we got a response rate of only 50 %. We also compared the response rate to three previous developmental study groups that used other scenarios. In Table 1, we see the ratio of interactions per group size for each of the groups and the average time spent interviewing the VH patient.

Table 1. Comparing the number of interactions and average time spent between the experimental group (with persuasion strategies), control group, and four prior developmental study groups

	Experimental	**Control Group**	**Previous Group**	**GI Hemm.**	**Post-Op Hemm.**	**STD**
Interactions/ Users	15 / 11	1 / 9	12 / 22	9 / 22	1 / 22	2 / 22
% responded	**81.82 %** [9/11]	11.11% [1/9]	50 % [11/22]	40.91% [9/22]	4.54% [1/22]	9.09% [2/22]
Average time	10.12 min	-	9.84 min	-	-	-

User participation: One-way ANOVA analysis of the number of interactions per user showed that there was a significant difference ($p < .001$) between at least two of the groups. Tukey HSD analysis showed that there were significantly more interactions per user for the *experimental group* than any of the five other groups ($p < .01$ for all groups). Thus, we accept **hypothesis-1**. Additional evidence supporting hypothesis-1 is that two out of the eleven experimental group participants recorded multiple interactions while only one out of the 88 users from all the four previous study groups had multiple interactions with the VH. Amongst the experimental group participants, one user interacted five times and another user interacted three times with the VH. A further reinforcement of this result is that the experimental and control groups were requested to participate at the end of a clinical rotation, the period in which they have the greatest demands on their time.

The behavior of the experimental group participants indicated that the "You were beaten by user [username]" e-mails had a significant effect. There were several times during the study where a participant would beat the score of their competitors, and this would cause a cascade of interactions. As an example we highlight an interaction on the fourth day after the recruitment e-mails were sent out. In this interaction a participant received a higher score than seven other users, causing a "beaten-by" e-mail to be sent to those seven users. Within eight minutes of the e-mail, three additional interactions occurred.

Duration of Interaction: The difference in average time between the experimental group and the previous group did not reach statistical significance (p=0.435). The users did not spend significantly more time with the VH in a single interaction. Therefore, the results of this study do not support **hypothesis-2**.

Difference in user's approach to the interaction: While analyzing the data, it was observed that participants from the experimental group behaved less realistically during the interaction. The realism of the interactions was evaluated through the use of greeting phrases. Nine out of eleven participants (82%) in the *previous group* used a greeting such as "Hello", or "How are you?" while one out of nine participants (11%) in the *experimental group* used a greeting. Using a two-tailed Fisher test shows this to be extremely statistically significant at $p < .001$. This finding highlights a possible limitation of these external motivation strategies; the scoring mechanism may influence user behavior in the VH interview. The participant's primary motivation when influenced by these motivation strategies appears to have been to get a high score, rather than to conduct a realistic medical interview.

6 Conclusion

We proposed motivation strategies for persuading users to participate in the conversational modeling of VHs. The evaluation study results indicate that the motivational strategies were effective in increasing user participation in the conversational modeling process. However, the users seem to approach the externally motivated human-VH interactions in a less realistic way. These results allow us to conclude that external motivation strategies can be helpful in increasing user participation during virtual human development, but care must be taken to align the scoring mechanism with learning goals in the stage of virtual human deployment.

References

1. King, D., Delfabbro, P., Griffiths, M.: Video Game Structural Characteristics: A New Psychological Taxonomy. International Journal of Mental Health and Addiction 8(1), 90–106
2. Bickmore, T., Schulman, D., Yin, L.: Engagement vs. Deceit: Virtual Humans with Human Autobiographies. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 6–20. Springer, Heidelberg (2009)
3. Dickerson, R., et al.: Evaluating a script-based approach for simulating patient-doctor interaction. In: International Conference on Human-Computer Interface Advances for Modeling and Simulating (SIMCHI 2005), pp. 79–84 (2005)
4. Rossen, B., Lind, S., Lok, B.: Human-Centered Distributed Conversational Modeling: Efficient Modeling of Robust Virtual Human Conversations. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 474–481. Springer, Heidelberg (2009)
5. Charon, R.: Narrative medicine: form, function, and ethics. Annals of Internal Medicine 134(1), 83 (2001)