Designing Embodied Conversational Agents to Conduct Longitudinal Health Interviews

Laura M. Pfeifer, Timothy Bickmore

Northeastern University College of Computer and Information Science
202 WVH, 260 Huntington Avenue, Boston, MA 02115
{laurap, bickmore}@ccs.neu.edu

1 Introduction

Embodied conversational Agents (ECAs) encompass many features that make them desirable as interviewers, specifically, as collectors of self-reported data. Cassell & Miller have proposed using ECAs as a mechanism for further understanding self-report behaviors, since they possess the features of a computer, providing a sense of anonymity to the interviewee, along with the features of a human, providing the ability to adjust, re-explain, and use non-verbal cues during the course of the interview [1]. One barrier to using ECAs to study self-report behaviors in-depth, is that most ECA systems are designed for short-term use, often just a single interaction. However, many nuances with self-report via an interview, such as social desirability, rapport and trust, unfold over time. Furthermore, there are specific applications where repeated, long-term interviews are necessary, e.g. for the daily discussion and monitoring of chronic disease.

The design of ECAs for such long-term use is a relatively understudied area, both within the HCI and Intelligent Agents communities. In this work, we discuss an ECA system designed to conduct automated health interviews, repeatedly, over long periods of time. We focus on factors that influence patient engagement with the system, and factors that influence both the quantity and quality of the data obtained.

2 Initial Longitudinal Study and Results

We have completed an initial longitudinal study designed to gather self-reported health information. In this study, we examined a system where participants interacted with an animated 2D ECA designed to emulate an exercise counselor, on their home computers for 12 weeks. At the end of each session, they had the opportunity to answer a single-item question: the number of minutes that they went for a walk. Participants were randomized into one of two conditions: self-reporting to the ECA, or self-reporting via a standard web form. We compared self-reported results to results gathered via a pedometer worn by participants throughout study.

During the first six-weeks of the study, participants in the Web group were more likely to self-report their minutes of walking, compared to those in the ECA group. However, self-reports given by participants in the ECA group were more accurate.

For the last six weeks of the study, conditions were flipped: participants that had been self-reporting to the ECA were now self-reporting via a web form, and vice-versa. During this time, participants in both groups increased the numbers of days that they interacted with the system, but reduced the number of times they voluntarily self-reported their walking behavior. Accuracy of the self-reported behavior improved over time, for both groups. Results are shown in Table 1.

Table 1. Study results, displayed by condition and study weeks. *** $p < 0.0001$

	Group 1 n = 13		Group 2 n = 12	
	Weeks 1-6 ECA	Weeks 7-12 Web	Weeks 1-6 Web	Weeks 7-12 ECA
% of days with a system interaction	60.4	65.8	55.8	68.3
% of system interaction days where self-reported data was given	74.2	69.4	95.0	91.3
Correlation between steps and minutes, Pearson's ρ	0.75***	0.84***	0.50***	0.56***

3 Ongoing Work

Based on our initial findings, we are conducting a longitudinal study to explore and manipulate several additional factors that could influence long-term system use and data quality. The first is **disclosure**: if the ECA informs the user *why* they are being asked to self-report their data, or given information regarding *how* the data will be used, or relayed the *importance* of the question being asked, will this effect the likelihood that someone will volunteer their information to the ECA, and will it influence the quality of that self-reported data? The second factor is **feedback**: if users are given feedback regarding the information they self-report to the agent, does this influence their system use later on? Furthermore, are there different effects between *generic* feedback and *personally relevant* feedback?

We will be conducting another randomized longitudinal experiment examining these factors in detail, within the context of automated, self-reported health data collection.

References

1. Cassell, J., Miller, P.: Is it Self-Administration if the Computer Gives you Encouraging Looks? In: Conrad, F.G., Schober, M.F. (eds.) Envisioning the Survey Interview of the Future, pp. 161-178. John Wiley & Sons (2007)