

Towards the Design of Multimodal Interfaces for Handheld Conversational Characters

Timothy Bickmore

MIT Media Lab

20 Ames St., Room E15-320

Cambridge, MA 02139

+1 617 253 7368

bickmore@media.mit.edu

ABSTRACT

This paper presents a study of individuals having conversations with animated characters on PDAs, and characterizes their use of natural nonverbal behavior compared to behavior exhibited in similar conversations with another person. The study finds that most people use the same nonverbal behavior in conversations with handheld characters that they use in conversations with people, although the frequency is somewhat lower in the handheld case. These results can inform the design of new PDA input modalities which leverage the natural nonverbal behavior observed.

Keywords

Embodied conversational agent, conversational characters, multimodal input, handheld computers, PDAs.

INTRODUCTION

There is a growing interest in using embodied conversational agents [2] as interfaces for mobile devices. Eyematic, Pulse, Anthropics, and Ananova Ltd. are currently developing animated character interfaces for use on PDAs. However, with the exception of Oviatt and Adams' study of speech disfluencies [5], little is known about the kinds of verbal and nonverbal behavior people would use when interacting with handheld embodied conversational agents (HECAs).

Although the results of empirical studies on the benefits of conversational agents have been mixed [4], at least one study has shown that users consider agents who use appropriate conversational behavior (gaze, turn-taking and limited hand gesture) to provide a significantly smoother interaction than agents which do not use these behaviors [3]. Studies of human-human interaction have also demonstrated that people pay attention to nonverbal behavior in conversation as separate channels of information from speech [1], and thus provide an

opportunity for building more robust and intuitive multimodal interfaces.

METHOD

A simulated HECA was constructed for the experiment and operated via a wizard-of-oz control. The "PDA" was actually a 3" color handheld television with a cover fabricated to hide the fact that it was a television (see Figure 1), connected to a computer workstation via a cable. The 3D character was capable of moving its lips in synchronization to synthetic speech, gesturing with its right hand, gazing either towards the user or a clipboard in its left hand, nodding its head, and rotating towards and away from the user. Subjects interacted with either the HECA or a human experimenter in a room equipped with two video cameras and a microphone to record their verbal and nonverbal behavior. The wizard sat in an adjacent room and controlled the character's responses via a computer.

Two structurally-identical health interview interaction scripts were created on the topics of exercise and healthy diets, based on publicly-available questionnaires augmented with open-ended questions designed to elicit a full range of conversational behavior (e.g., "describe a stairmaster machine to me in as much detail as you can"). The scripts began with a greeting and six turns of small talk, followed by the body of the question-and-answer session, two turns of wrap-up ("thank you for participating") and a farewell.

12 (7 male, 5 female, mostly students) subjects participated in a 2 condition (HUMAN vs. HECA), completely-balanced (with respect to topic and condition), within-subjects experiment.

At the start of each session, subjects were told they would be interacting with either a person or an "animated

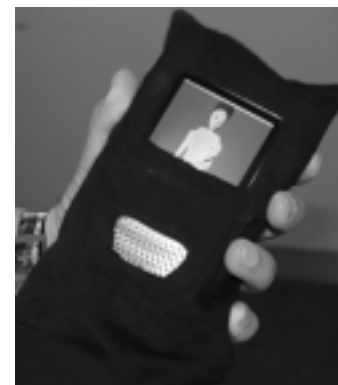


Figure 1. Simulated HECA

LEAVE BLANK THE LAST 2.5 cm (1") OF THE LEFT
COLUMN ON THE FIRST PAGE FOR THE
COPYRIGHT NOTICE.

	<i>HUMAN</i>			<i>HECA</i>		
	<i>Subjects who used</i>	<i>% Turns with</i>	<i>Average per turn</i>	<i>Subjects who used</i>	<i>% Turns with</i>	<i>Average per turn</i>
Hand Gestures	12	23%	0.91	9	8%	0.30
Gaze Lookaways	12	50%	0.97	11	32%	0.50
Head Nods	12	63%	1.53	11	26%	0.67
Eyebrow Raises	10	11%	0.19	8	11%	0.14

Table 1. Use of Nonverbal Behavior in Two Conditions

computer character" who "played the role of a health advisor who will ask you questions about the health aspects of your lifestyle, such as exercise and eating a healthy diet". In the HECA condition subjects were also told "you will interact with him the same way that you interact with another person, just by talking to him", at which point they were handed the PDA and the experimenter left the room.

Behavioral measures included: number of hand gestures per turn; speaker gaze lookaways per turn (away from and back to listener to hold the floor); headnods per turn (used for backchannel feedback, agreement, and emphasis); and eyebrow raises per turn (used for emphasis, to indicate request for feedback, or to indicate confusion).

RESULTS

The data yielded a corpus of 720 subject speaking turns of interaction. There was an average of 31 turns of interaction per session (min 27, max 34), lasting an average of 6 minutes and 51 seconds.

The frequency of nonverbal behavior observed in the two conditions is summarized in Table 1. All hand gestures occurred within the normal gesture space for each subject; subjects were never observed to hold their hands in front of the HECA or make tiny gestures in the HECA's apparent field of view. Of the 114 gestures produced by subjects interacting with the HECA, only 9 of the gestures involved the hand holding the PDA (see Figure 2, middle).

During a post-experiment debrief, 8 of the 12 subjects said they felt that the image of the animated character helped their interaction. Ten of the subjects reacted positively when asked if they liked talking to a PDA ("Like a phone conversation almost, except for the little guy.", "At first it was weird, then it seemed pretty natural and normal").

DISCUSSION

Overall, this study confirms that people use the same conversational modalities when talking with a handheld character as when talking face-to-face with another person.



Figure 2. Example Nonverbal Behavior: (L) Free hand gesture; (M) Two-handed gesture; (R) Gaze away

Taken together, these behaviors indicate that subjects were in fact engaging in a natural, multimodal face-to-face conversation with the HECA, and neither the size of the character nor the fact that they were holding it in their hands seemed to have a significant impact on their overall mode of interaction.

The presence of frequent gaze lookaways in the HECA condition indicates that gaze cues could provide an important input modality for mobile devices, since they can help determine when a user is speaking to the device rather than to other people nearby in addition to determining when a user is starting or finishing his or her turn. Most subjects also used headnods in their HECA conversations, and at least once per session a subject was observed to give a headnod-only response to a statement made by the HECA. A gaze-tracking system could also be used to provide these head motion cues. The presence of frequent hand gestures also motivates the development of tracking devices for the free hand, in order to use these gestures as an additional input modality.

ACKNOWLEDGMENTS

Thanks to Justine Cassell, Neil Sengupta, Tom Stocky, Yukiko Nakano, Hannes Vilhjálmsson and Kimiko Ryokai.

REFERENCES

1. Cassell, J., McNeill, D., and McCullough, K. Speech-Gesture Mismatches: Evidence for One Underlying Representation of Linguistic and Non-Linguistic Information. *Pragmatics and Cognition* 6, 2, 1998, 1-33.
2. Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.
3. Cassell, J. and Thorisson, K. The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents. *Applied Artificial Intelligence* 13, 1999, 519-538.
4. Dehn, D. and van Mulken, S. The Impact of Animated Interface Agents: A Review of Empirical Research. *International Journal of Human-Computer Studies* 52, 2000, 1-22.
5. Oviatt, S. and Adams, B. Designing and Evaluating Conversational Interfaces with Animated Characters. *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and Churchill, E. (eds.). MIT Press, Cambridge, MA, 2000, 319-345.