Officer Garcia: A Virtual Human for Mediating Eyewitness Identification

Brent Daugherty*
University of North Carolina at Charlotte

Sabarish Babu†
University of Iowa

Brian Cutler*
University of North Carolina at Charlotte

Larry Hodges*
University of North Carolina at Charlotte

Abstract

An analysis of court cases has revealed that the mistaken identification of the wrong person by victims and witnesses of a crime is the single most common error leading to the arrest and conviction of innocent people [Wells et al. 2006]. Recognizing the role of mistaken identification in erroneous conviction, a growing number of states and police departments have reformed their eyewitness identification procedures. In this paper, we investigate a new procedural reform: the use of a virtual officer who does not know the identity of the suspect in the lineup and therefore cannot bias the witness toward false identification.

CR Categories: K.4 Computers and Society, H.5 Information Interfaces and Presentation.

Keywords: Virtual Humans, Virtual Officer, Officer Garcia, Embodied Conversational Agents, Human-Computer Interaction

1 Introduction

Investigations by the Innocence Project (www.innocenceproject.org), using DNA evidence, have led to the exoneration of over 200 convicted felons, many of whom spent over ten years in prison, and some of whom were sentenced to death. Their research and prior archival studies have established that mistaken identification is the leading precursor of erroneous conviction [Wells et al. 2006].

In recognition of the role of mistaken identification in erroneous identification, some states and many police departments have revised their identification procedures, taking into account scientific research on factors affecting the suggestiveness of identification tests [Wells et al. 2006]. Among the procedural reforms is the use of "blind" administrative procedures. Blind procedures are those in which the investigator conducting the identification is unaware of the suspect's identity. By being blind to the suspect's identity, the investigator is unable to advertently or inadvertently influence the eyewitness's selection of the suspect.

Thus far, the states of New Jersey and North Carolina and police departments in various cities (e.g., Minneapolis, Boston) have adopted blind procedures together with additional reforms. One of the main concerns about blind procedures is their practicality. Some police departments have resisted this reform because they do not have the personnel to implement it. To address officers' influence and the limited availability of blind lineup administrators, research and commercial software have been introduced in the last few years to conduct the lineup autonomously using a standard window and mouse interface.

While promising, the standard computer interface has significant limitations. The stress associated with witnessing a crime compounded with that of learning a new computer program (or worse, an eyewitness that lacks experience with computers) renders the interface impracticable in some cases, though the magnitude of this problem is not currently known. According to one police officer with whom we conferred as part of this research, there exists another more subtle issue where eyewitnesses are sometimes not receptive to volunteering their time and energy when technology is involved (even recording devices have been inauspicious), whereas they do not display this uncooperative behavior while conversing with an officer about the incident.

This paper will discuss how a virtual human interface agent was constructed iteratively for administering a photographic lineup through guided conversation. There are several advantages to using a virtual human interface agent towards a social collaborative task such as a photographic lineup. Since virtual humans are modeled after humans, these interfaces can use several modalities for communicating information, such as gestures and facial expression, which are "transparent" to the user [Cassell 2000]. Furthermore, research evidence suggests that people can achieve tasks effectively when the behavior and attitude of an interface agent is similar to a real human [Takeuchi 1999]. These benefits come from the strength of the virtual human interface metaphor that allows us to leverage people's experience with real social interaction to enrich human-computer interaction.

Our contribution includes a preliminary user study with user feedback from four professors and two students of computer science, one professor (an expert in eyewitness memory) from psychology, and a local police sergeant. The lessons learned from the iterations also serve as an informal methodology for virtual human construction and directing future research.

2 Related Work

Few computerized eyewitness identification tools are currently available. PC_Eyewitness, developed by Maclin et al. [2005], using Visual Basic, has been the only one of the computerized tools subjected to user evaluation and compared with paper-and-pencil – the de facto standard for comparing computerized tools in psychological labs. However, another program, Remote Lineup (http://www.sungard.com/hte), is currently in development in the private sector for implementation by some North Carolina police departments [Barksdale 2007] and for use in some forthcoming

*e-mail:{bsdaughe, blcutler, lfhodges}@uncc.edu

†e-mail: sabarish-babu@uiowa.edu

VRST 2007, Newport Beach, California, November 5–7, 2007.
© 2007 ACM 978-1-59593-863-3/07/0011 $5.00

national field studies to be conducted under the auspices of the American Judicature Society's Institute for Forensic Science and Public Policy. However, these tools do not make use of interactions beyond the WIMP (Windows, Icons, Mouse, Pointer) interface.

Researchers have shown that a virtual human interface can provide feedback to individuals using a variety of verbal and non-verbal behaviors. Zanbaka et al. [2004] found that people respond to virtual humans similarly to the way they respond to real humans, and anecdotal evidence suggests that human communication consists of a high bandwidth of several modalities such as gestures, facial expressions, speech, and body language [Cassell 2000].

Virtual humans have already been used in training. Slater et al. [2000], examined the effectiveness of a virtual human-aided rehearsal for a live performance and found that a performance level was reached that could not be achieved by only learning of lines or video conferencing. To improve the doctor-patient relationship, virtual humans have allowed medical students to practice their interpersonal skills [Raij et al. 2006]. Finally, Babu et. al. [2007] showed that immersive virtual humans in natural, multi-modal interaction can be used as a tool for teaching users social conversational protocols in a foreign culture.

Our contribution is in the use of a virtual human interface agent for eyewitness identification through guided conversation and interactive feedback. Our virtual officer was constructed iteratively though evaluation and feedback from several subject matter experts.

3 Virtual Officer Framework & Iterations

The Virtual Officer Framework (VOF) is built on top of the Virtual Human Interface Framework (VHIF) [Babu et al. 2005] to combine cutting edge speech and graphics with an extensible discourse modeler and planner to guide the eyewitness through the identification procedure. Officer Garcia is able to interact using a combination of spoken natural language with non-verbal cues that include maintaining eye contact, facial expressions, and gestures, as well as maintaining human-like social communication protocols such as turn-taking, feedback, and repair mechanisms.

In the following iterations, we discover many concerns of the VOF and the research questions they present. Concerns occur in the form of subtle, virtual human inconsistencies which are important to address due to the sensitive nature of this kind of application – and thus may make the system unusable – as well as higher level comments that advocate replacing computer graphics and speech synthesis with their human-recorded counterparts.

3.1 Iteration One

The VOF allows the virtual officer's gender, race, and language to be interchanged. The importance of officer manipulation is apparent in, for instance, rape cases (use a same-gender officer) or non-native speakers (speak in their native language). However, for the purpose of this preliminary evaluation, the virtual officer and language are not changed.

The non-verbal cues are pre-scripted using key-frame techniques afforded by the Haptek (www.hapek.com) motion generation engine. The motion generation engine is also used to evoke the appropriate non-verbal behavior based on the scripted content.

Figure 1: Iteration One. Officer Garcia gesturing to a photo during the identification procedure.

Haptek Lip Synch Engine was used for lip synchronization. Dragon Naturally Speaking 7.3 was used for speech recognition with AT&T Text-to-Speech for speech utterance. See [Babu et al. 2005] for a detailed description.

Officer Garcia is outfitted with a typical navy blue police uniform, including a badge, nameplate, two side patches, and a tie. The background resembles a police office with a projection screen that is utilized for the photographic lineup, as illustrated in Figure 1.

The guided conversation closely follows the procedures used in a large, Southeast metropolitan police department. Specifically, Officer Garcia's instructions conform to the North Carolina Actual Innocence Commission Recommendations for Eyewitness Identification (http://www.ncids.org/News & Updates/Eyewitness ID.pdf) with dichotomous questioning, e.g., yes or no, so as to minimize misrecognition.

The VOF also allows an officer to upload photos where Officer Garcia will display them sequentially, rather than simultaneously – a recommended practice for reducing misidentifications due to comparative identifications. After each recognized photograph, the virtual officer records an audio file of why the victim recognized the photograph and their confidence in the selection. The complete transcript and audio files are logged.

3.2 Iteration Two

The next iteration is a result of feedback from two students and four professors from the Department of Computer Science and one professor from the Department of Psychology who is an expert in eyewitness memory – all from UNC Charlotte. Since this iteration consisted of domain experts, we opted for first-impressions (instead of directed questioning):

- How will gestures be interpreted? Could these gestures be misconstrued as biased?
- What if the user wants to go back in the interaction or a speech command was incorrectly recognized?
- What if the eyewitness is hysterical? Will the virtual human still be effective?
- Is a windows and mouse interface a fair comparison? Maybe touch screens, such as those used for testing at the [North Carolina] Department of Motor Vehicles, would be a more significant research question?

Figure 2: Iteration Two. On the left, Officer Garcia leaves the viewing area before the photographic lineup is presented (on the right). The "Help" command remains on screen throughout the procedure.

- They also identified certain words or phrases that were not easily heard and could be misinterpreted. (Single-syllable words caused the most problems.)

The claim that a virtual officer could have a bias was surprisingly the most common comment on the system – the very complaint this system was suppose to address. Nevertheless, there are many random gestures programmed into the virtual officer to make him appear more lifelike, and since a quick blink of the eye could be noticed during one photo but not on the others, there exists a possibility that the officer projects a bias.

As a result of this endowed bias and to make the interaction for each suspect consistent, any movement not associated with breathing or talking was removed. During the most influential part, the photo lineup, the virtual officer completely leaves the viewing area and the screen is focused on the photographs (see Figure 2).

The next issue raised was option selection and error correction. Restricting the eyewitness to two voice commands was not enough to compensate for microphone noise and foreign accents. So now the user must confirm the selection before the VOF will continue. This has the additional benefit of requiring the eyewitness to think a moment longer about his or her decision.

A virtual officer's ability to deal with a hysterical witness is a realistic concern considering eyewitness lineups are typically reserved for serious crimes. This anxiety has been alleviated in three ways: (1) requiring confirmation to selections, (2) easing into the virtual human interface through small talk as well as an explanation and example of the process, and (3) incorporating the voice command, "Help," to signal the need for a real officer to return.

Finally the issue was raised that a mouse, keyboard, and window interface is not a fair comparison for a virtual human. Certainly a touch screen is only one option; there are many modes of input that are possible. The most apt solution to this problem is to compare the virtual officer with a real officer who is blind to the identity of the suspect.

3.3 Iteration Three

The last iteration is based on the recommendations of a local police sergeant with little background using computers. He completed an evaluation and discussed with us his experience and concerns with eyewitness interviews:

- The virtual officer's speech [synthesis] is choppy [compared to human speech].
- A virtual human will not be taken seriously. It would be better to record a T.V. reporter-like investigator [instead of the virtual human] and then show the eyewitness the pre-recorded video.
- I wasn't sure how many photos would be shown. Would it continue until I identified a photo?
- The instructions could be more concise, but the length was fine. (About 8 minutes.)
- The introduction was crucial for getting through the procedure.
- The commands should stand out, and the background is distracting.
- Uniformed police officers don't necessarily conduct the eyewitness investigation, e.g. a police uniform is unnecessary.

The problem with the virtual officer's irregular speech could be solved by using human-recorded speech as opposed to speech synthesis. This is possible due to the extensible nature of the framework, in which the virtual officer is capable of lip synching to audio files. Similarly, using video of real humans would certainly add the missing sense of realism and seriousness.

The verboseness of the procedure is unavoidable since the virtual human is leading the conversation and the script must abide by the recommendations of the NC Actual Innocence Commission. However, he did confirm that the introduction was necessary to complet the procedure. So, though the police sergeant thought the script could be more concise, the opening dialog is important for creating rapport and understanding how to interact.

The remainder of his suggestions, as shown in Figure 3, were aesthetic: emphasize the commands, remove distractions (office scene), and change the outfit of the virtual officer.

Despite being unable to recruit more police officers for this iteration, there are compelling reasons to think his views are in the majority. Specifically, his superior rank (sergeant) and several years of experience conducting and teaching (officers) how to conduct eyewitness interviews speaks to his credibility. In addition, several discussions have since reiterated his concerns.

Figure 3: Iteration Three. Officer Garcia is now wearing a suit with the projection screen covering the background. On the right, the user is asked, "Do you recognize this person?" The lineup number is indicated on each photo.

4 Summary and Future Work

The movement to reform eyewitness identification procedures nationwide is gaining momentum. Several critical limitations have been identified, such as the lack of personnel to administer blind lineups, and new, untested technological solutions. Toward this end, computer-based lineups have many advantages. However, computers pose yet additional limitations in that they may surpass the experience of some crime witnesses and may therefore be impractical in actual investigations. The virtual officer, using natural methods of communication, is likely to reduce the difficulties currently associated with WIMP interfaces.

In this research we discuss the design and implementation of a virtual officer agent (Officer Garcia) for assisting in suspect identification using natural, multi-modal interaction through guided conversation. Officer Garcia was iteratively constructed using evaluation and feedback from subject matter experts, which has elucidated several research questions:

- Does a virtual human add any more bias?
- How do people perceive him?
- Do they trust him and the system?
- How does the bias, perception, and trust of a pre-recorded investigator compare with a virtual human?
- How does the bias, perception, and trust of a live investigator compare with a virtual human?

The comments from the preceding iterations have directed the VOF toward two future evaluation studies using crime-simulation methodology: College students will view videotaped enactments of crimes and will be randomly assigned to, in the first study, a virtual officer or live investigator and, in the second study, a virtual officer or pre-recorded video of an investigator, to attempt identifications from a photo lineup. Questionnaires containing rating scales and open-ended questions will assess the following: (1) basic demographic characteristics (age, gender, ethnicity, and occupational status) and (2) experience with crime and computers; (3) positive and negative affect of the identification procedure; (4) comfort with and ease of use of the identification procedure; (5) ratings of and suggested improvements to the virtual officer as well as (6) a co-presence questionnaire to determine the extent they felt that they were interacting with another human.

References

BABU, S., SUMA, E., BARNES, T., HODGES, L.F. 2007. "Can Immersive Virtual Humans teach Social Conversational Protocols?" In *Proceedings of Virtual Reality 2007*, IEEE.

BABU, S., SCHMUGGE, S., INUGALA, R., RAO, S., BARNES, T., AND HODGES, L. F. 2005. Marve: a prototype virtual human interface framework for studying human-virtual human interaction. In Lecture Notes in Computer Science, T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, Eds. Springer-Verlag, London, 120-133.

BARKSDALE, T. 2007 Police to Change Lineups. Winston-Salem Journal. Retrieved June 12, 2007. http://www.journalnow.com/servlet/Satellite?pagename=WSJ%2FMGArticle%2FWSJ_BasicArticle&c=MGArticle&cid=1149192559485

MACLIN, H. O., ZIMMERMAN, A. L., AND MALPASS, S. R. 2005. PC Eyewitness and the Sequential Superiority Effect: Computer-Based Lineup Administration. *Law and Human Behavior*, Vol. 29, No. 3, June 2005.

RAIJ, A., J., K., DICKERSON, R., LOK, B., COHEN, M., STEVENS, A., BERNARD, T., OXENDINE, C., WAGNER, P., AND LIND, D. S. 2006. Interpersonal Scenarios: Virtual ≈ Real? In *Proceedings of Virtual Reality 2006*, IEEE, 59-66.

SLATER, M., HOWELL, J., STEED, A., PERTAUB, D-P., GARAU, M., AND SPRINGEL, S. 2000. Acting in Virtual Reality. *ACM Collaborative Virtual Environments 2000*, pp. 103-110, 2000.

TAKEUCHI, Y. AND KATAGIRI, Y. 1999. Social Character Design for Animated Agents. *RO-MAN99*.

WELLS, G. L., MEMON, A., & PENROD, S. D. 2006. Eyewitness Evidence: Proving its Probative Value. *Psychological Science in the Public Interest*, 7, 45-75.

ZANBAKA, C., ULINSKI, A., GOOLKASIAN, P., HODGES, L. F. 2004. Effects of Virtual Human Presence on Task Performance. *Proceeding of the International Conference on Artificial Reality and Telexistence (ICAT)*, pp. 174-181.