Animated Virtual Agents to Cue User Attention

Comparison of static and dynamic deictic cues on gaze and touch responses

Santiago Martinez, Robin J.S. Sloan, Andrea Szymkowiak and Ken Scott-Brown
University of Abertay Dundee
Dundee, DD1 1HG. UK
s.martinez@abertay.ac.uk, r.sloan@abertay.ac.uk, a.szymkowiak@abertay.ac.uk, k.scott-brown@abertay.ac.uk

Abstract — **This paper describes an experiment developed to study the performance of virtual agent animated cues within digital interfaces. Increasingly, agents are used in virtual environments as part of the branding process and to guide user interaction. However, the level of agent detail required to establish and enhance efficient allocation of attention remains unclear. Although complex agent motion is now possible, it is costly to implement and so should only be routinely implemented if a clear benefit can be shown. Previous methods of assessing the effect of gaze-cueing as a solution to scene complexity have relied principally on two-dimensional static scenes and manual peripheral inputs. Two experiments were run to address the question of agent cues on human-computer interfaces. Both experiments measured the efficiency of agent cues analyzing participant responses either by gaze or by touch respectively. In the first experiment, an eye-movement recorder was used to directly assess the immediate overt allocation of attention by capturing the participant's eye-fixations following presentation of a cueing stimulus. We found that a fully animated agent could speed up user interaction with the interface. When user attention was directed using a fully animated agent cue, users responded 35% faster when compared with stepped 2-image agent cues, and 42% faster when compared with a static 1-image cue. The second experiment recorded participant responses on a touch screen using same agent cues. Analysis of touch inputs confirmed the results of gaze-experiment, where fully animated agent made shortest time response with a slight decrease on the time difference comparisons. Responses to fully animated agent were 17% and 20% faster when compared with 2-image and 1-image cue severally. These results inform techniques aimed at engaging users' attention in complex scenes such as computer games and digital transactions within public or social interaction contexts by demonstrating the benefits of dynamic gaze and head cueing directly on the users' eye movements and touch responses.**

Keywords-agents; digital interface; touch interface; computer animation; reaction time; eyetracking.

I. INTRODUCTION

The allocation of attention by a human observer is a critical yet ubiquitous aspect of human behaviour. For the designer of human-computer interfaces, the efficient allocation of user attention is critical to the uptake and continued use of their interface designs. Historically, many human-computer interfaces have relied on static textual or pictorial cues, or a very limited sequence of frames loosely interconnected over time (for example, on automated teller device menus, or on websites). More recently, the increased power of computer graphics at more cost effective prices has allowed for the introduction of high resolution motion graphics in human computer interfaces. Until now, psychological insights on attention and the associated cognitive processes have mirrored Human-Computer Interaction's (HCI) reliance on either static or stepped pictorial stimuli, where stepped pictorial stimuli consist of a few static frames displayed over time to imply basic motion. Again, this legacy can be attributed to limitations in affordable and deployable computer graphics.

This paper extends previous work from CONTENT 2010 Martinez et al. [1] and is centered on the evaluation of fully animated (25 frames per second) virtual agents, where both the head and eye-movements of the agent are animated to allocate user attention. In contrast to most previous studies that have relied on manual inputs, using peripheral devices in response to agent cues, this research explores the possibility of two different ways of interaction. The first study uses the captured eye-gaze of participants as a response mechanism, following on from the work of Ware and Mikaelian [2], while the second study explores the suitability of attention allocation involving small amount/range of locomotion (i.e., touch action) on the same task.

Where observers look in any given scene is determined primarily by where information critical to the observer's next action is likely to be found. The visual system can easily be directed to guide and inform the motor system during the execution of information searching. Consequently, a record of the path that observer gaze takes during a task provides researchers with what amounts to a running commentary on the changing information requirements of the motor system as the task unfolds [3]. This is the underlying principle of the reported experiment, which is an expansion of the cognitive ethology concept expressed by Eastwood et al. [4] to virtual agents. The experiments are based on the deictic gaze cue – the concept that the gaze of others acts like a signal that is subconsciously interpreted by an observer's brain, and that it can transmit "information on the world" [5]. The gaze of another human agent is inherently difficult to avoid, and it can be used as a specific pointer to direct an observer's attention [6]. The incorporation of this concept can be easily implemented into an agent-based interface.

Another aspect this study evaluates is how locomotion can influence the effectiveness of cueing. Most research has been focused on response using peripheral devices. It is important to assess the validity of cues on a wider range of modalities. In this study we analyze gaze and touch inputs in order to assess the suitability of agent cues in these kind of interfaces and their applicability in upcoming interface design.

The efficiency of interfaces such as these can be assessed based on the speed of observer response to cues. In both studies, the cues are presented as fully animated (dynamic) agents, stepped agents (two images), or static agent images (one image). Coupled with appropriate software, a virtual agent can anticipate a user's goals, and point (using gaze) to the area where the next action has to be performed. An agent with animated gaze may therefore be useful to adopt in digital interfaces to guide user attention and potentially increase the speed of attention allocation, or where the work space of human physical action may have many possible choices; and the possibility of not selecting the right one is high.

In the following sections, we will explain in detail the application of the virtual agent to cue user attention. In Section 2 we will describe the existing literature reviews from two different research fields. In Section 3, we will explain the method used to develop a gaze experiment. Gaze input results will be presented in Section 4. In Section 5, we will describe the method of a touch experiment and in Section 6 its results. Finally, in Section 7, we will discuss the overall conclusions of both experiments: dynamic versus static cues, the differences observed between the interaction modalities (e.g., gaze and touch) and the impact on user engagement and agent animation on real world interfaces.

II. LITERATURE REVIEW

Previous studies belong to two different but related research fields: namely cognitive psychology and computer interface design. Psychological studies have reviewed attention and its relationship with the cues. Posner [7] describes the process of orienting attention. Relative to neutral cue trials, participants were faster and/or more accurate at detecting a target given a valid cue, and they were slower and/or less accurate given an invalid cue. Friesen and Kingstone [8] worked with faces and lines drawn following the gaze direction towards the target area. They found that subjects were faster to respond when gaze was directed towards the intended target. This effect was reliable for three different types of target response: detection, localization and identification. Langton and Bruce [9], and more recently Langton et al. [10], investigated the case of attention in natural scene viewing. They concluded that facial stimuli, that indicate direction by virtue of their head and eye position, produce a reflexive orienting response in the observer. Eastwood et al. [4] produced experimental findings leading to the conclusion that facial stimuli are perceived even when observers are unaware of the stimuli. In 2006, Smilek et al. [11] focused on isolating specific processes underlying everyday cognitive failures. They developed a measure for attention-related cognitive failures with some success, and introduced the term of cognitive ethology. Studies in HCI and computing are mostly focused on proving the validity of eye-gaze as an input channel for machine control. One exception was Peters et al. [12] in 2009, who tested shared attention behaviours during virtual agent interaction. The method was based on a head direction mapping metric (directedness) using their own algorithm and recorded by a common and cheap available equipment, a webcam. They demonstrated, with some success, the importance of participant head motion directed to an object in the interface to infer the level of engagement. However, the absence of gaze tracking disabled the analysis of peripheral eye movements and covert attention. Also, the use of a gaze-contingent moving cross-hair was an important distractor on the tasks, becoming intrusive.

Concerning the study of the eye-gaze as an input modality, Ware and Mikaelian [2] used an eye-tracker to compare the efficacy of gaze with other more usual inputs, such as manual using physical devices. They found that the gaze input was faster with a sufficient size of target. Sibert and Jacob [13] studied the effectiveness of eye gaze in object selection using their own algorithm and compared gaze selection with a traditional input – a hand operated mouse. They found that gaze selection was 60% faster than mouse selection. They concluded that the eye-gaze interaction is convenient in workspaces where the hands are busy and another input channel is required.

The above research shows how eye-gaze can be used to assess the response of a user when accurate tracking is possible. In addition, it has been demonstrated that the eye-gaze of an agent can effectively allocate attention. However, the interplay between pictorial cues to gaze allocated attention (and subsequent assessment of allocated attention) is still to be fully explored. Specifically regarding this point, for the reported experiment, two goals were set by the authors; to assess the extent to which the gaze of the observer can be used to record their selection of targets and response time to agent cues, and to determine whether fully animated agents would offer an advantage over standard static (1-image) or stepped (2-image basic motion) agents when directing attention using gaze. By focusing on gaze as a means of target selection, this removes as much motor response as possible from the observer. Manual responses operated through any device inevitably introduce uncertainty in establishing the true response time since they are an indirect response to the gaze cue (requiring over allocation of attention and eye-gaze, followed by translation of the response signal to the input modality of device). Therefore, when it comes to assessing the effectiveness of animated versus static and stepped agent cues, directly recording the eye-movements of observers and using this data to determine the speed of their response and their selection of objects offers a significant advantage.

Nevertheless, touch inputs are increasingly appearing in our daily lives on screens, via smartphones, kiosks or Automated Teller Machines (ATMs). In this context, touch is considered a natural way of interaction [14]. It rapidly evolved from places where there was no space for peripherals (i.e., factory environment), such mouse or

keyboard, to be included in portable devices, desktop PCs, and home entertainment.

Originally, use of touch screens was limited by a lack of precision, high error rates of selections [15] and absence of ergonomic standards in their physical design. Although touch calibrations are still required in some devices such as eye-trackers and large touch-sensitive display screens, the evolution of screen technologies (i.e., capacitive, resistive, surface acoustic wave [16]) has largely solved the precision and errors in selection problems. At the same time, advances in hardware design have tackled the ergonomic standards problem by allowing the user to adjust screen position by independent rotations on two axes for fixed monitors and three axes for tablet screens. Touch screen applications are beginning to be found in many different contexts, such as information kiosks, airports, education, museums, amusement parks, and very widely on self-service technologies (e.g. Schreder et al. cite the railways usage [17]). Attributes of touch screens are: fast response time [14] (especially in most recent generation of hardware), contribution to user satisfaction [18] [19] and above all direct manipulation of elements (an important advantage for infrequent users of interfaces [20]).

Previous research states that directly touching the screen provides a more direct approach to elements on the interface, conferring a more natural way of handling objects than with a mouse or other pointer device [21]. Ever since Jef Hann's TED talk [22] (which has since clocked up over 5million views), the repertoire of touch screen modality has been evolving towards a standard for user input. Wobbrock (2009) outlines a multitude of gestures available for the Microsoft Surface, but in the meantime a more reduced selection of gestures is becoming apparent through the development of hardware and associated applications. This is most obvious in the form of multi-touch mass produced items such as iPod Touch and more recently iPad (with 15 million sales at time of writing). The sales of touch screen computers are testament to the engaging qualities of the interface possibilities. However, how the direct physical action affects the interface elements' performance compared to interactions with external buttons, track pads, trackballs or mice remains uncertain. In the context of public space touch screens, this uncertainty comes from two areas. Firstly from the influence of layout: when there are no peripheral buttons required to be used (e.g. keyboard or mouse) the keys get you around an ATM [18]. The second uncertainty comes from the use of fingers to touch the screen, fingers and hands that can occlude large parts of the screen and thus change the layout requirements of a display. In this work, we evaluate whether the touch task constrains or interferes to some extent with how cues allocate the attention of user and whether these agent cues are as effective on a touch screen as in a non-contact, gaze, interface.

Figure 1: The appearance of the virtual agent, surrounded by eight target squares, arranged on both the cardinal and oblique axes.

III. GAZE-RESPONSE EXPERIMENT: METHOD

The experiment method was as follows below.

A. *Task description*

In this experiment, participants were asked to perform an object selection task (using their eye gaze alone) on a series of twenty-four different agent animations, presented on a monitor at a resolution of 1024 x 768. Each of the videos showed a virtual agent's head in the centre of the screen surrounded by eight different possible target areas (see Fig. 1). Each agent was displayed on screen for 3000 ms. Over the course of the video, the agent would orient its head and eyes to aim at a particular target square. The point at which the agent oriented its head and gaze (and the nature of the agent's movement) was determined by the type of agent cue (see below). Of the eight target areas in each video, only one was the right choice in each trial – the one that was specifically indicated by the agent. If the participants selected that specific area with their eye-gaze, it was counted as a success. If the participant selected any of the other seven areas, it was counted as incorrect. Fixations to areas outside the 8 target areas were coded as no target selected. The target areas were red squares approximately 150 x 150 pixels in size, and were all equidistant from the center of the screen.

B. *Agent Cues*

There were three different types of agent cues (see Fig. 2):

a) Static cue: A single image of an agent. The agent's head and eyes were aimed at the target area for the duration that the stimulus is displayed. The orientation cue was therefore presented from 0 ms till 3000 ms.

b) Stepped cue: Two images of an agent, sequenced to imply movement. The agent's head and eyes were looking straight forward from 0 ms, before the second image was displayed from 960 ms. In the second image, the agent's head and eyes were aimed at the target from 960 ms till 3000 ms.

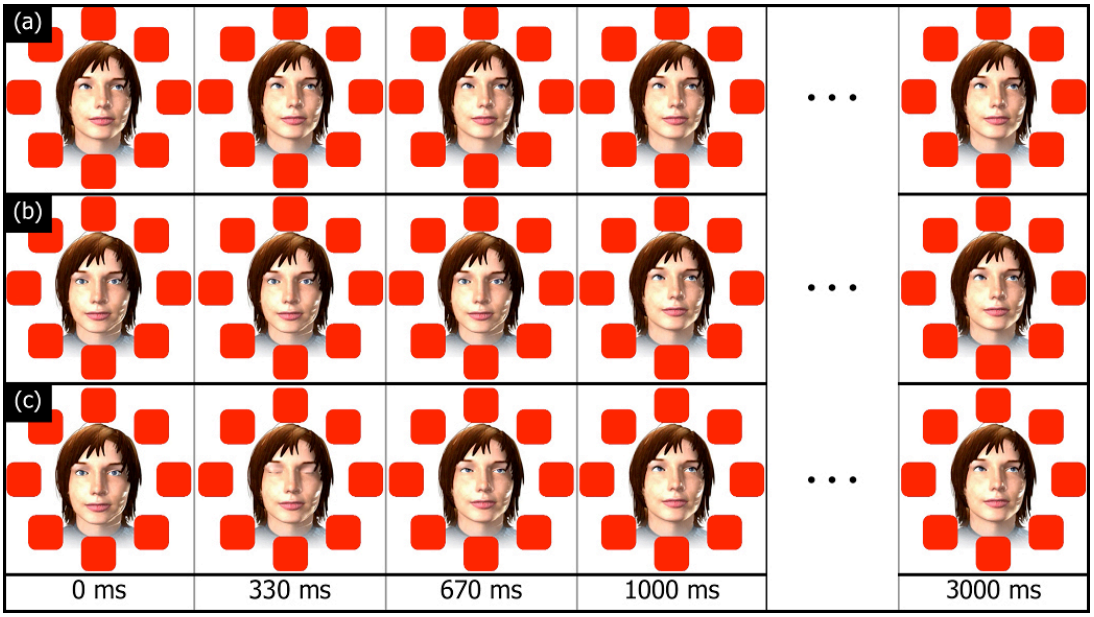

Figure 2. The appearance of the three types of helper agents over 1000 ms. Helper agents used head orientation and gaze to highlight one of eight targets. In the above example, three types of helper agent are shown highlighting the NE target. (a) shows a static (1-image) helper agent, which highlights the NE target from 0 ms onwards. (b) shows the stepped (2-image) helper agent, which looks towards the observer in frame 1 (from 0 ms) before changing to highlight the NE target in frame 2 (from 960 ms). (c) shows the dynamic (25-image, 25 fps) agent, which begins at 0 ms by looking at the observer, and is animated with natural movement so that the head and gaze shift towards the NE target at 960 ms. All helper agents are shown. to participants for a total of 3000 ms, so that the appearance of the agent at 1000 ms is held for two seconds.

c) Dynamic cue: A fully animated agent, showing naturalistic movement from 0 ms to 960 ms. The agent's head and eyes were pointing straight forward at 0 ms, before the agent moved (at 25 fps) to aim its head and eyes at the target area. The agent's gaze and head were aimed at the target at 960 ms. The full orientation cue was therefore presented from 960 ms till 3000 ms.

C. Participants

A total of sixteen participants were recruited from students and staff at the University of Abertay-Dundee. There was no compensation and all had normal or corrected-to-normal vision. During the experiment, two of them used contact lenses.

D. Apparatus

To capture participant gaze data, a modified (fixed position) SMI IView HED eye-movement recorder with two cameras was used. One camera recorded the environment (the target monitor) and the other tracked the participant's eye by an infrared light recording at a frequency of 50 Hz and accuracy of 0.5° of visual angle. Stimuli were presented on a TFT 19'' monitor with a 1024 x 768 resolution and 60Hz of frequency controlled by a separate PC. The monitor brightness and contrast were set up to 60% and 65% respectively to ease the cameras' recordings and avoid unnecessary reflections. In addition, both devices were individually connected to two different computers. Viewing was conducted at a distance of 0.8 meters in a quiet experimental chamber.

Each participant underwent gaze calibration controlled by the experimenter prior to the start of data collection. The participant was sat down in a height adjustable chair with their chin on the chin rest and in front of the monitor at 0.9 meters distance. Firstly, the calibration of the eyetracker was completed by presenting a sequence of five separate dots in the center and in each of the corners. The calibration covered the same surface occupied by the target areas.

A final image with the set of five points was shown to double check the calibration by the operator. The calibration was repeated if necessary following adjustments to the camera positions to ensure good calibration. The experiment started with a ten seconds countdown sequence. After that, the series of twenty-four videos (3 agent cue types x 8 target areas) were presented to participants in a randomized order. The duration of each task video was three seconds, and each video was shown one by one in full screen mode. Before each task video, a central black cross over a white background was shown for two seconds to center the gaze of the participant. This ensured that the participant was looking at the centre of the screen at the start of each video. Fig. 3 shows sample screen captures from the eye-tracker.

E. Data analysis

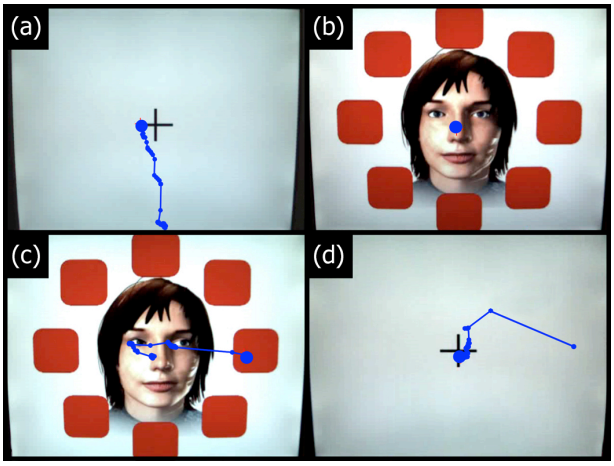

Figure 3: The eye tracking data of one participant, where the blue circles represent fixations. In image (a), the participant looks towards the cross before the agent appears in image (b). In image (c), the agent highlights the East target, at which point the participant looks towards this target, before fixating on the cross again in image (d).

The participant gaze data was analyzed using the software BeGaze 2.3. The data stored in BeGaze contained all the fixations' timestamps. Only trials where the participant's gaze started on the cross in the center of the screen were considered valid. Target selection was defined by the first full-gaze fixation occurring in the eight predefined areas of interest overlying the 8 target destinations. The fixation duration criterion for an observer response is defined in the light of previous literature. Ware and Mikaelian in 1987 used 400 ms; Sibert and Jacob in 2000 considered 150 ms. Because extended forced fixation (400 ms) can become laborious, we established a criterion for successful cognitive response to fixation as equal as or greater than to 250 ms, i.e., a fixation that locates on the target area at least for 250 ms.

Based on this concept, of the total number of possible cognitive responses, 92.18% were successfully tracked. Of the successfully tracked data, correct responses accounted for 95.2% of the total and mismatches accounted for 4.9%. The definition of a mismatch was when there was a fixation of 250 ms or more inside an incorrect target area. In 8.47% of the total mismatches, no clear target was selected – i.e., there was no fixation of 250 ms or more in any of the target areas.

IV. GAZE-RESPONSE EXPERIMENT: RESULTS

Only one participant presented problems during the tracking because of the unexpected movement of her contact lens in the tracked eye. This resulted in four non-tracked responses in the same participant.

For each agent type a total of 128 eye tracking recordings were made. Recordings were then evaluated and allocated to one of four categories: Correct (where the observer clearly selected the intended target), Incorrect (where the observer clearly selected an unintended target), No Target (where it was not clear which target the observer had selected), and Corrupted (where the eye tracking data had been disrupted resulting in lost data, for instance when a participant's head moved in a trial). After excluding the corrupted recordings, it was clear that observers were able to accurately select the intended target regardless of whether the virtual agent was static (95%), stepped (92.5%), or dynamic (94.2%) (see Table I). This would suggest that, in general, the type of virtual agent (in terms whether it was static, stepped, or fully animated) did not substantially impact upon how effective it was at communicating what the intended target was.

TABLE I. PARTICIPANT GAZE SELECTION OF TARGETS

Type	Correct	Incorrect	No Target selected	Corrupt (Exclusions)
Static	93.5 %	5.7 %	0.8 %	7 / 128
Stepped	92.5 %	5.8 %	1.7 %	8 / 128
Dynamic	94.2 %	5 %	0.8 %	7 / 128

A repeated measures analysis of variance (ANOVA) was used to determine whether agent type had an effect on how long it took participants to look at and select the intended target square. The response times for static agent cues - which contained agents that were oriented towards the target 960 ms earlier than both stepped and dynamic cues – were corrected to account for this difference. The analysis showed that the type of agent did have a significant effect on participant response time, $F(2, 30) = 52.73$, $p < .001$.

Participants responded most quickly to the dynamic (fully animated) agent type (M = 1220, SE 95) than they did to either the stepped (2 frame) agent type (M = 1874, SE 61) or the static (1 frame) agent type (M = 2091, SE 59) (see Fig. 4).

Comparisons between agent types were assessed using a Bonferroni post-hoc test. The results showed that participants responded to the dynamic agent type significantly more quickly than both the static (Mean Deviation (MD) = 870, $p < .001$) and the stepped (MD = 654, $p < .005$) agent types. Furthermore, participants also responded to the stepped agent type significantly more quickly than the static agent type

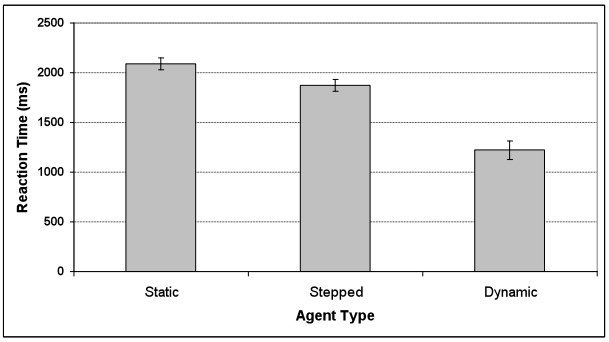

Figure 4: The mean gaze response times for static, stepped, and dynamic agents indicate that participants reacted most quickly to the fully animated, dynamic agents

TABLE II. MULTIPLE COMPARISON BETWENN AGENT TYPES (GAZE)

Type	Comparison	Mean Deviation	Std. Error	Sig.
Static	Stepped	217 ms	54.3	.004
	Dynamic	870 ms	85.8	.000
Stepped	Static	-217 ms	54.3	.004
	Dynamic	654 ms	114.3	.000
Dynamic	Static	-870 ms	85,8	.000
	Stepped	-654 ms	114.3	.000

(MD = 217, $p < .005$) (see Table II). These results not onlyunderline that static agent types are significantly less effective at cueing observer attention than either stepped or dynamic agents, but also that stepped agent types are significantly less effective than fully animated, dynamic agents.

V. TOUCH-RESPONSE EXPERIMENT: METHOD

The experiment method was as follows below.

A. Task description

The task to perform in this experiment was analogous to the described above (see Section 3 *Gaze-Response Experiment*), except this time hand-touch was the input modality instead of eye-gaze. Participants had to perform the object selection task using the same hand for all trials, on a series of twenty-four different agent animations. Agent animations were presented on a monitor at a resolution of 1280 x 720. Each of the videos showed a virtual agent's head in the centre of the screen, surrounded by eight different touchable square areas (see Fig. 5). Each agent was displayed on screen for 3000 ms and remained on the screen with the last frame shown till a target selection was made by participant. Orientation cues timing, type of agents and type of choices are identical as previously described in gaze experiment.

Over the course of the video, the agent would orient its

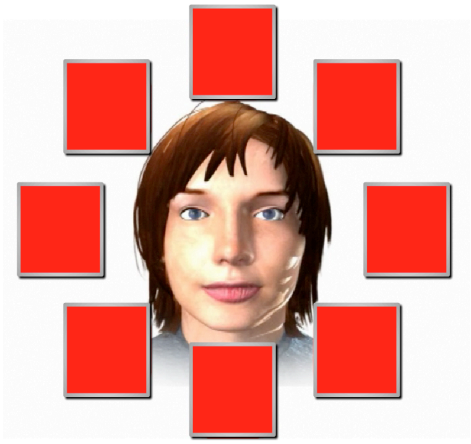

Figure 5: The appearance of the virtual agent, surrounded by eight target squares, arranged on both the cardinal and oblique axes.

head and eyes aim at a particular target square. The point at which the agent oriented its head and gaze (and the nature of the agent's movement) was determined by the type of agent cue. Of the eight possible target areas in each video, only one was the right choice in each trial – the one that was specifically indicated by the agent. If the participants selected that specific area, it was counted as a success. If the participant selected any of the other seven areas, it was counted as incorrect.

The target areas were red squares of exactly 150 x 150 pixels in size, and were all equidistant from the center of the screen. In comparison with gaze experiment, targets have the same area but with the slight difference in the layout, a grey border around the border to create a button similarity –giving a 'push-able' notion to the eight square items.

B. Agent Cues

The agent cues described in section III.B were also used in the current experiment.

C. Participants

A total of thirty-two participants were recruited from students and staff at the University of Abertay-Dundee. 4 participants already participated in the gaze experiment. There was no compensation and all had normal or corrected-to-normal vision and were able to use hands correctly for the purpose of this experiment. There were 29 right-handed, 1 left handed and 2 ambidextrous. Both ambidextrous participants chose right hand to run the experiment. In one case choice was the participant's dominant-hand and in the other it was their non-dominant hand. They were asked to use the same hand across all trials and all participants did so.

D. Apparatus

The trials were run in a Sony VAIO® L Series Touchscreen. It is an All-In-One PC multi-touch (two point) capacity on the screen (dimensions of 24 inches at 60 Hz; resolution of 1280x720; bright and contrast at 62%, graphic card default levels). Computer specifications were memory of 4 GB DDR2 SDRAM, processor Intel® Core™ 2 Duo CPU E7500@, 2.93 GHz and 2.94 GHZ. The OS was Microsoft® Windows 7 Home Premium 64 bits. The video card was an integrated GeForce G210M with a total graphics memory of 2271 MB (512 MB dedicated). The PC was securely placed on an office table and participants were seated on a chair with adjustable height. The PC was at a distance of approximately 25 cm from participant's head and 12 cm from participant's hands, well within arm's reach. All the trials were run in a quiet chamber in the Usability Lab of the University of Abertay-Dundee. During the experimental trials, the experimenter was observing the experiment in a separate twin room through a one-way mirror to minimize the disturbance or possible noises on participants. Participants were told they could raise their hand in any moment to request presence of the researcher. During the thirty-two runs, the researcher's assistance was required only once due to equipment failure. All of this participant's trials were removed from the analysis and all their response data discarded.

After the participant was comfortably sat in the chair and contented with the distance of PC, button-feedback training was run to make them confident with the button touch feedback. It was recommended to make at least one touch per target area (n = eight) to feel how the buttons worked. The experiment started with a ten second countdown video. Before each trial, a text indicated to participant to push *space bar* key to start. This assured that the participant was resting their hand at the same point before the start of each trial. The series of twenty-four videos (3 agent cue types x 8 target areas) were presented to participants in a randomized order and counter-balanced. The duration of each video was three seconds, and each video was shown one by one. Preceding each stimulus trial, a central black cross over a white background was shown for two seconds (similarly as seen in Fig. 3.a) to mirror gaze experiment task preamble. The last video frame from each trial remained on screen until the participant selected one of the eight touchable areas.

E. Data analysis

The participant response time data was stored using Adobe Flash CS5 (version 11.0.0.485). The data contained all the participants time responses (24 per participant) counted from the starting point of showed cue (video with the agent) till the participant selected a target area on the screen by their finger touch. Successful target selection was defined by the touch on the target area cued specifically by the agent. A touch in any of the other seven areas not cued by the agent was considered a mismatch. No responses outside the eight target areas were recorded during the experimental trials.

The choice of Adobe Flash to measure Reaction Time (RT) was intentional. First, it gave a desired degree of freedom in the design of the interface layout. In contrast with gaze experiment where all elements on the interface where passive, here the touchable areas or buttons are external to the video and now become functional components themselves. Second, the decision was based on studies proving the validity of Flash as reliable software to measure RT, once that specific conditions were accomplished in the experiment. One condition is related with the device used in the time measurement, in this case the PC. The smaller the difference in RTs, the more critical it is to know the properties of the timing device used. Neath et al. [23] showed that the smallest difference in magnitude that a stock iMac 8.1 (April, 2008-March, 2009) using Flash (version 10.0 r42) could detect under realistic conditions is approximately 5–10 ms, and this dictates the types of research that should use these systems: if a researcher tests all subjects using the exact same hardware, if the focus is on relative rather than absolute RTs, if the differences in RTs in the conditions to be examined are expected to be fairly large (e.g., at least 20–40 ms), if only certain software is used, and if many properties of the visual display are not of critical importance, then the conclusions drawn from RT data collected on a stock iMac are likely to be the same as those drawn from RT data collected on custom or high-end hardware.

Reimers et al. [24] in 2007 studied on PC (processor 1.4 MHz AMD Athlon, 256 MB of RAM, graphic card PCI NVidia GeForce 2MX and 32 MB of video RAM) the estimation of the average and the spread of RTs in the different conditions stating that RTs recorded with Flash are between 10 and 40 ms longer than those recorded in the Baseline condition (application on programming language C using the X Window System to display stimuli and a parallel port button box). Flash did not appear to add significant random error to RT measurements.

VI. TOUCH-RESPONSE EXPERIMENT: RESULTS

An unexpected operating system error resulted in data loss of one participant, due to a sudden failure of OS that invalidated the participant's session. All of the participant's trials were removed from the analysis and all his times discarded. Thus, 94.8% of the total number of responses were successfully stored. Of these stored answers, correct responses accounted for 98.48% and mismatches accounted for 1.51%.

For each agent type a total of 24 time response recordings were made per participant. Recordings were then analyzed and allocated to one of three categories: Correct (where the observer selected the intended target), Incorrect (where the observer selected an unintended target), and Corrupted (where the file writing was corrupted or non-existent). After excluding the corrupted recordings, it was clear that users were able to accurately select the intended target regardless of whether the virtual agent was static (99%), stepped (99.7%), or dynamic (99.7%) (see Table III). This would suggest that, analogously as in the previous case of gaze-interaction, the type of virtual agent (in terms whether it was static, stepped, or fully animated) did not substantially impact upon how effective it was at communicating what the intended target was.

A repeated measures analysis of variance (ANOVA) was used to determine whether agent type had an effect on how long it took participants to select by touch the intended target square. The response times for static agent cues - which contained agents oriented towards the target 960 ms earlier than both stepped and dynamic cues – were corrected to account for this difference. The analysis showed that the type of agent did have a significant effect on participant response time, $F(2, 724) = 50.38$, $p < .001$. Participants responded most quickly to the dynamic (fully animated) agent type (M

TABLE III. PARTICIPANT TOUCH SELECTION OF TARGETS

Type	Correct	Incorrect	Corrupt (Exclusions)
Static	99%	1 %	24 / 256
Stepped	99.7%	0.3%	24 / 256
Dynamic	99.7%	0.3%	24 / 256

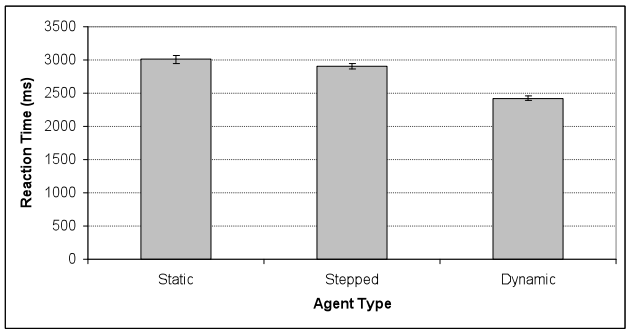

Figure 6: The mean touch response times for static, stepped, and dynamic agents indicate that participants reacted most quickly to the fully animated, dynamic agents

= 2423, SE 32) than they did to either the stepped (2 frame) agent type (M = 2900, SE 40) or the static (1 frame) agent type (M = 3007, SE 57) (see Fig. 6).

Comparisons between agent types were assessed using the Bonferroni post-hoc test. The results showed that participants responded to the dynamic agent type significantly more quickly than both the static (Mean

Deviation (MD) = 584, $p < .001$) and the stepped (MD = 476, $p < .001$) agent types. In contrast to gaze case, participants responded to the stepped agent type not significantly faster than the static agent type (MD = 108, $p > .005$) (see Table IV). These results corroborate the gaze experiment results where static agent types are significantly less effective at cueing observer attention than dynamic agents, but also that stepped agent types are significantly less effective than fully animated, dynamic agents.

VII. GAZE- AND TOUCH-RESPONSE EXPERIMENT: DISCUSSION AND FUTURE WORK

Using a paradigm where the criterion for correct response to pictorial or animated agent gaze is the eye-gaze of the participant we found that the presence of full-motion in the gaze and head inducing agent drives the observer's attention the fastest. Gaze recorded responses for 25 frame stimuli were 35% faster than stepped and 42% faster than static stimuli. This result is consistent with previous research on gaze cueing [10]. The current paradigm provides the most direct route to the establishment of the overt allocation of gaze location since it subverts the need for a translation to a device manual response. This confirms Ware and Mikaelian's [2] assertion that participants eye-gaze itself can be used to indicate responses.

By modifying the gaze cue paradigm from experiment to a touch-based target selection paradigm, we demonstrated that fully animated agent gaze and head cues drive user attention faster than static and 2-image agent cues. Touch recorded responses for 25 frame stimuli were 17% faster than stepped and 20% faster than static stimuli, confirming those obtained in eye-gaze interface experiments. Compared with the gaze response results, the decrease on the time differences suggests that the touch selection method alters, to some extent, the delay from when the participant correctly

TABLE IV. MULTIPLE COMPARISONS BETWEEN AGENT TYPES (TOUCH)

Type	Comparison	Mean Deviation	Std. Error	Sig.
Static	Stepped	108 ms	62.1	.249
	Dynamic	584 ms	62.1	.000
Stepped	Static	-108 ms	62.1	.249
	Dynamic	476 ms	61.6	.000
Dynamic	Static	-584 ms	62.1	.000
	Stepped	-476 ms	61.6	.000

follows the cue to when the target selection is executed. It seems that the motor response reduces the time advantages gained with the fastest cue, suggesting that eye responses are much more rapid than hand responses. There is an aspect, clearly observed, of longer reaction times in touch modality, probably due to the translation of response into the sense of touch. This fact reinforces the idea of the complex process of motor response that reduces the time saved by the motion cue. Such a process should be greater in magnitude in order to explain those time save absorptions. Confirming that the introduction of hand locomotion does not invalidate the effectiveness of dynamic cue, results also showed that it was the difference on time response between 2-stepped and static was non-significant.

Regarding whether the 2-image agent could be considered not completely a motion cue, this suggests that motion cues with a sufficient number of frames (25 tested in the experiment) are more necessary in context where human locomotion is involved. Probably the presence of touch involves more factors than those that we could control and include in the study, but at least one of them should be the higher impermeability to attention cues. The presence of movement in gaze cueing stimuli seems to drive the user's attention more quickly. One prediction arising from this is that, when compared with 2D agents, 3D agents create an expectation of more believable behaviour. The combination of additional pictorial cues and natural motion may make the appearance of the agent more akin to that of a human conversation partner. The additional realism possible with modern computer animation techniques may make agents more believable and engaging [25].

The present study indicates how the animation of an agent can be linked to the sequencing of the social 'script' or 'narrative' of a HCI interface experience. Previous investigators such as Kendon [26] observed a hierarchy of body movements in human speakers; while the head and hands tend to move during each sentence, shifts in the trunk and lower limbs occur primarily at topic shifts. They discovered the body and its movements as an additional part of the communication, participating in the timing and meaning of the dialogue. Argyle and Cook [27] discuss the use of deictic gaze in human conversation. They argued that during a conversation the eye gaze serves for information seeking, to send signals and to control the flow of the conversation. They explained how listeners look at the speaker to supplement the auditory information. Speakers on

the other hand spend much less time looking at the listener, partially because they need to attend to planning and do not want to load their senses while doing so. Preliminary work from our laboratory suggests that experience in the gaze task over time may lead to a learning effect whereby extended exposure to these stimuli leads to improved gaze allocation. This analysis will form part of a wider study including a sequence of guided navigation prompts in a naturalistic setting. Only by creating a natural sequence of user choices with a combination of gaze cues and items competing for attention (including distractors) can we fully confirm the efficacy of an agent-based cue in human computer transactions in the natural environment.

The research presented here is consistent with the wider conclusions of other investigators [25], which indicate that vivid, animated emotional cues may be used as a tool to motivate and engage users of computers, when navigating complex interfaces. The results of this experiment provide guidance for agent design in consumer electronics such as computer games or animation. In order to avoid an unpleasant robotic awareness, natural motion and the correct presentation of the cue contribute to increase the deictic believability of the agent. Deictic believability in animated agents requires design that considers the physical properties of the environment where the transaction occurs. The agent design must take account of the positions of elements in and around the interface. The agent's relative location with respect to these objects, as well as social rules known from daily life, are critical to create deictic gestures, motions, and speech that are both effective, efficient and unambiguous. All these aspects have an effect in addition to the core the response time measure. They easily trigger natural and social interaction of human users, reaching the right level of expectations. Furthermore, they make the system errors, human mistakes and interaction barriers more acceptable and navigable to the user [28].

Fully animated agents have the potential to be a key new component into the assistive characteristic of interfaces, where an appropriate animated performance demonstrating a solution to a problem can be delivered. In principle, this study has demonstrated that agent guidance would be suitable both with gaze and touch interfaces, but its use could be extendable to general interfaces, where searching and selection tasks are dominant. Animated agents could become a new component in the salience characteristic of interfaces, where a synchronized movement can reinforce the perceptibility of relevant elements inside the interface.

The concept of natural interfaces has been extensively discussed in the HCI literature [29]. The 'naturalness' is explained in terms of more familiarity, intuitive and predictable use, information retrieval and behaviour of the interface and the machine. In this context, the findings of the current study could be used to propose that more human-like interface components would be of practical use, particularly with agent behaviour synchronized with cues. The combination has a potential role in attention conflict situations, influencing significantly the overt allocation of user attention and, consequently, his responses and the interaction in general. In considering the graphical fidelity of agents, it is worth noting that natural realism can cause problems within interface design. Research examining expression animation by Zamitto et al. [30], highlights a valuable distinction between realism and believability. One of their conclusions was that the pursuit of realism in expression animation may risk falling into the 'uncanny valley' where increasing photo-realism results in a perception of falseness [31]. In the context of user interaction, we predict that such prioritization of realism on the agent could result in an inappropriate user disengagement. Instead, believability, suitability of the context and usefulness in their assigned task (i.e., timing regards the cues) should be the premises for the representation of the interface agent.

In future work, it is planned to extend this study with a wide range of emotions on the agent cues, to evaluate their suitability on these interfaces, and compare with the results already obtained with fully animated non-emotional agents. With these set of studies, it is intended to draw a better and more complete picture of new ways of implementing guidance in interface design. This guidance strategy attempts to cover the 'what to do next?' situation for new or infrequent users, and it is specifically designed to resolve attention conflicts on environments with many distractors in number and type, such as those typically found in public space interaction.

As an ultimate goal, this and future related work pursues the intention to effectively design methods of allocation of the attention of users to improve the interaction flow. It is crucial that we evaluate the cues used in a guidance system based on the principle of cueing that can anticipate user actions and help in 'what to do next' problems.

ACKNOWLEDGMENT

Martinez is grateful for support from the Alison Armstrong Studentship. Sloan is grateful for support from a University funded studentship. The authors gratefully acknowledge the supportive interdisciplinary research environment provided by Abertay's Whitespace Research Group, Institute for Arts Media and Computer Games and Centre for Psychology.

REFERENCES

[1] S. Martinez, R. J. S. Sloan, A. Szymkowiak, and K. C. Scott-Brown, "Using virtual agents to cue observer attention," in CONTENT 2010, The Second International Conference on Creative Content Technologies, 21-26 November 2010, Lisbon, Portugal, 2010.

[2] C. Ware and H. H. Mikaelian, "An evaluation of an eye tracker as a device for computer input", SIGCHI Bull., 17, May. 1986, pp. 183-188, doi:10.1145/30851.275627.

[3] J. M. Findlay and I. D. Gilchrist, "Active vision: the psychology of looking and seeing", Oxford University Press, Oxford. 2003.S. R. H.

[4] J. D. Eastwood, D. Smilek, and P. M. Merikle, "Differential attentional guidance by unattended faces expressing positive and negative emotion", Perception & Psychophysics, vol. 63, 2001, pp. 1004-1013.

[5] I. Poggi and C. Pelachaud, "Signals and meanings of gaze in animated faces,". In: S. Nuallain, C. Muhlvihill and P.

McKevitt, eds, Language, Vision and Music. Amsterdam: John Benjamins, 2001

[6] S. R. H. Langton and V. Bruce, "Reflexive visual orienting in response to the social attention of others", Visual Cognition, vol. 6, 1999, pp. 541-567., doi:10.1080/135062899394939

[7] M. I. Posner, "Orienting of attention", Quarterly Journal of Experimental Psychology, vol. 32, 1980, pp. 3–25.

[8] C. K. Friesen and A. Kingstone, "The eyes have it! reflexive orienting is triggered by nonpredictive gaze", Psychonomic Bulletin and Review, vol. 5, 1998, pp. 490-495.

[9] S. R. Langton, R. J. Watt, and V. Bruce, "Do the eyes have it? Cues to the direction of social attention", Attention And Performance, vol. 4(2), 2000, pp. 50-59, ISSN 1364-6613, DOI: 10.1016/S1364-6613(99)01436-9

[10] S. R. Langton, C. O'Donnell, D. M. Riby, and C. J. Ballantyne, "Gaze cues influence the allocation of attention in natural scene viewing", Experimental Psychology, vol. 59(12), 2006, pp. 2056- 2064, doi: 10.1080/17470210600917884.

[11] D. Smilek, E. Birmingham, D. Cameron, W. Bischof, and A. Kingstone, "Cognitive ethology and exploring attention in real world scenes," Brain Research, vol. 1080, Issue 1, Attention, Awareness, and the Brain in Conscious Experience, 2006, pp. 101–119.

[12] C. Peters, S. Asteriadis, and K. Karpouzis, "Investigating shared attention with a virtual agent using a gaze-based interface", Journal on Multimodal User Interfaces, vol. 3, Dec. 2009, pp. 119-130.

[13] L. E. Sibert and R. J. Jacob, "Evaluation of eye gaze interaction", In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 00), ACM, 2000, pp. 281-288, doi: 10.1145/332040.332445.

[14] A. Sears, "High precision touchscreens: design strategies and comparisons with a mouse", International Journal of Man-Machine Studies, vol. 34, Apr. 1991, pp. 593-613.

[15] J. Pickering, "Touch-sensitive screens: the technologies and their application", International Journal of Man-Machine Studies, vol. 25, Sep. 1986, pp. 249-269.

[16] M. Platshon, "Acoustic touch technology adds a new input dimension", Computer Design, Mar. 1988, pp 89-93.

[17] G. Schreder, K. Siebenhandl, E. Mayr, & M. Smuc, "The ticket machine challenge? Social inclusion by barrier-free ticket vending machines". In Proceedings of the The good, the bad and the challenging: The user and the future of information and communication technologies (pp. 780-790), 2009

[18] N. P. Marcous, M. J. Brant, and M. J. Rosenzweig, System and method for electronic transfer of funds using an automated teller machine to dispense the transferred funds. Google Patents, 1997.

[19] M.D. Stone, "Touch-Screens for Intuitive Input", PC Magazine, 1986, 183-192.

[20] R.L. Potter, L.J. Weldon, and B. Shneiderman, "Improving the accuracy of touch screens: an experimental evaluation of three strategies", Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '88, 1988, pp. 27-32.

[21] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan, "Direct-touch vs. mouse input for tabletop displays", Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07, 2007, p. 647.

[22] J. Hann's TED talk http://www.ted.com/talks/jeff_han_demos_his_breakthrough_touchscreen.html <retrieved: June, 2011>

[23] I. Neath, A. Earle, D. Hallett, and A.M. Surprenant, "Response time accuracy in Apple Macintosh computers", Behavior research methods, Mar. 2011, pp. 1-10-10.

[24] S. Reimers and N. Stewart, "Adobe Flash as a medium for online experimentation: A test of reaction time measurement capabilities", Behavior Research Methods, vol. 39, Aug. 2007, pp. 365-370.

[25] T. Vanhala, V. Surakka, H. Siirtola, K. Raiha, B. Morel, and L. Ach, "Virtual proximity and facial expressions of computer agents regulate human emotions and attention", Computer Animation And Virtual Worlds, vol 21(3-4), 2010, pp. 215-224, doi: 10.1002/cav.336.

[26] A. Kendon, "Some relationships between body motion and speech: ana anlysis of an example", In: A. Siegman and B. Pope, eds, Studies in Dyadic Communication, pp. 177-210, Elmsfor, NY: Pergamon Press, 1972.

[27] M. Argyle and M. Cook, "Gaze and mutual gaze", New York: Cambridge University Press, 1976, 221 pages, ISBN-13: 978-0521208659.

[28] E. M. Diederiks, "Buddies in a box: animated characters in consumer electronics",IUI '03, 2003, pp. 34-38, doi: http://doi.acm.org/10.1145/604045.604055.

[29] W. Buxton, "Lexical and pragmatic considerations of input structures", ACM SIGGRAPH Computer Graphics, vol. 17, no. 1, p. 31–37, 1983.

[30] V. Zammitto, S. DiPaola, and A. Arya, 2008, "A methodology for incorporating personality modeling in believable game characters", Arya, 1(613.520), p.2600.

[31] M. Mori, "The uncanny valley", Energy, vol. 7, no. 4, p. 33–35, 1970.