Believability Testing and Bayesian Imitation in Interactive Computer Games

Bernard Gorman[1] Christian Thurau[2] Christian Bauckhage[3] Mark Humphrys[1]

[1] Dublin City University, Glasnevin, Dublin 9, Ireland
{bgorman,humphrys}@computing.dcu.ie
[2] Bielefeld University, D-33501 Bielefeld, Germany
cthurau@techfak.uni-bielefeld.de
[3] Deutsche Telekom AG, 10587 Berlin, Germany
christian.bauckhage@telekom.de

Abstract. In imitation learning, agents are trained to carry out certain actions by examining a demonstration of the task at hand. Though common in robotics, little work has been done in translating these concepts to computer games. Given that present-day games generally use antiquated AI techniques which can often lead to stilted, mechanical and conspicuously artificial behaviour, it seems likely that approaches based on the imitation of human players may produce agents which convey a more humanlike impression than their traditional counterparts. At the same time, there exists no formal method of quantifying what *constitutes* a 'humanlike' impression; an equivalent of the Turing test is needed, with the requirement that an agent's appearance and behaviour be capable of deceiving an observer into misidentifying it as human. The aims of this paper are thus threefold; we describe an approach to the imitation of strategic behaviour and motion, propose a formal method of quantifying the degree to which different agents are perceived as 'humanlike', and present the results of a series of experiments using these two systems.

1 Introduction

Imitation learning, as the name suggests, refers to the acquisition of skills or behaviors through examination of a demonstrator's execution of a given task. Imitative techniques have been adopted by many researchers in robotics as a means of 'bootstrapping' their machines' intelligence, providing them with a high level of competence after a comparatively short training period [1]. Demiris and Hayes [2], for instance, train an apprentice robot to navigate a maze by imitating the actions of a demonstrator agent. Schaal [3] proposes a control-based approach to imitating a tennis swing from demonstration. Fod, Mataric and Jenkins [4] outline various statistical approaches to deriving movement primitives from observed human motion.

Despite the interest exhibited by the robotics community, however, very few attempts have been made to apply these principles to interactive computer games. Indeed, even the most modern games still predominantly rely on symbolic artificial intelligence techniques that were developed several decades ago [5, 6]. Given that many modern games allow the recording of entire sessions, and that – rather than limb movement data, as is common in robotic imitation – these recordings encode the frame-by-frame behaviour of the player under complex, rapidly-changing conditions and in competition with opponents of comparative skill, it becomes clear that computer games are an ideal platform for research in imitation learning. In this paper, we detail part of our work in this area; a Bayesian-based approach to the derivation and imitation of human strategic behaviour and motion patterns in commercial computer games. In conjunction with the believability-testing system described below, we then demonstrate its effectiveness in producing convincingly humanlike game agents.

When evaluating imitation agents, three distinct metrics are applicable: i) **statistical** analysis of the accuracy with which the observed behaviours are reproduced; ii) **believability testing** to verify whether the cloned agent effectively conveys the impression of being human; iii) **performance-based** assessment of the imitation agent in direct competition against other agents and human players. This paper concerns itself with believability testing. A significant impediment to work in this field is the lack of a formal, rigorous standard for determining how 'humanlike' an artificial agent is, or any strict means of comparing the believability of different agents. While some contributions compare observers' reactions to artificial and human players [7], these have invariably been of a very limited, informal and often inconclusive nature. Imitation learning holds obvious potential as a method of producing more credible agents, but there has thus far been no means of empirically assessing this credibility; the need for a perception- and behavior-based analogue to the Turing test is clear [8]. To address this need, we introduce a formal method of quantifying the degree to which cognitive agents are perceived as 'humanlike', and of facilitating the objective comparison of different agents. This method has been designed to minimize the subjectivity associated with such surveys, and to produce a *believability index* weighted according to both the observer's experience and the certainty with which the agents are identified.

The *first-person shooter* (FPS) genre – wherein players explore a 3D environment littered with weapons, bonus items, traps and pitfalls, with the objective of defeating as many opponents as possible – was chosen for our work on the basis that it provides a relatively *direct* mapping of human decisions onto agent actions. Due to its prominence within the literature [9], we opted to use iD Software's Quake II® as our testbed. In order to extract the required data from its recorded *DM2* or *demo* file format – consisting of the network traffic received during the game - and to realise the in-game agents (or *bots*, in game vernacular), we employ our own QASE API and its MatLab-integration facilities [10].

2 Imitation Learning - Methodology

In this section, we outline our current approach to imitating human movement and strategic behavior in Quake II® . The individual components of this ap-

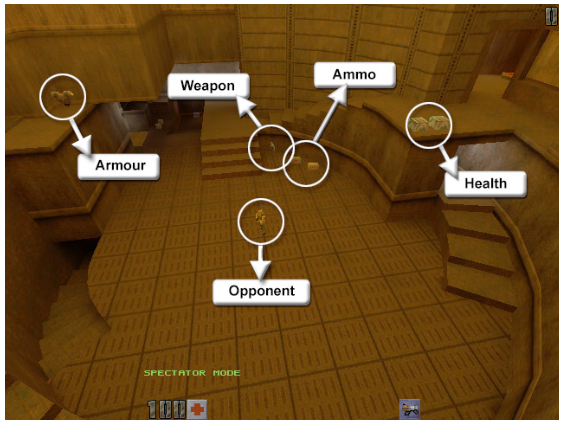

Fig. 1. Typical QUAKE II® environment

proach were introduced in previous publications [11, 12], while a forthcoming contribution describes their integration. Here, we briefly review the system; readers are referred to the earlier publications for additional details.

2.1 Behaviour Model

The current model focusses on two core aspects of human behaviour; *strategic planning* and *motion modelling*. A number of investigations [13, 7] have found that the ability of an agent to exhibit long-term strategic planning faculties is a crucial factor in determining how humanlike it appears. The importance of *motion modelling* is equally evident - human players frequently exhibit actions other than simply moving along the environment surface, including jumps, weapon changes and discharges, crouches, etc. In many cases, the player can only attain certain goals by performing one or more such actions; they therefore have an important *functional* element. From the perspective of creating a believable agent, it is also vital to reproduce the *aesthetic* qualities they encode.

2.2 Learning Goal-Oriented Strategic Behaviours

In QUAKE II® , experienced players traverse the environment methodically, controlling important areas of the map and collecting *items* to strengthen their character. Thus, we define the player's long-term goals to be the items scattered at fixed points around each level. By learning the mappings between the player's status and his subsequent item pickups, the agent can adopt observed strategies when appropriate, and *adapt* to situations which the player did not face.

We first read the set of all player locations $\boldsymbol{l} = [x, y, z]$ from the recording, and cluster the points using a *fast k-means* to produce a *goal-oriented* discrimination of the level's topology. We also construct an $n \times n$ matrix of edges E, where

n is the number of clusters, and $E_{i,j} = 1$ if the player was observed to move from node i to node j and 0 otherwise. The player's *inventory* – the list of what quantities of which items he currently possesses – is also read from the demo at each timestep, and unique state vectors are obtained; these *inventory prototypes* represent the varying situations faced by the player during a game. We can now construct a set of *paths* which the player followed while in each such situation.

Having obtained the different paths pursued by the player in each inventory state, we turn to reinforcement learning to learn his behaviour. The topological map of the game environment may now be viewed as a *Markov Decision Process*, with the clusters corresponding to states and the edges to transitions.In this scenario, the MDP's actions are considered to be the choice to move to a given node from the current position. Thus, the transition probabilities are $P(c' = j|c = i, a = j) = E_{ij}$ where c is the current node, c' is the next node, a is the executed action, and E is the edge matrix. We assign an increasing reward to consecutive nodes in every path taken under each prototype, such that the agent will be guided along similar paths to the human when facing similar situations. With the transition probabilities and rewards in place, we now run a modified version of the *value iteration algorithm* in order to compute the utility values for each node in the topological map under each inventory state prototype.

A number of other features of human planning behaviour must also be taken into account. Principal among these are the human player's intuitive *weighing* of strategic objectives, and his understanding of *object transience* – that is, a collected item will be unavailable until the game regenerates it after a fixed interval. To model these, we introduce a weighted *fuzzy clustering* approach and an *item activation* variable:

$$m_p(\boldsymbol{s}) = \frac{a(o_p)d^{-1}(\boldsymbol{s}, \boldsymbol{p})}{\sum a(o_i)d^{-1}(\boldsymbol{s}, \boldsymbol{i})} \tag{1}$$

where m is the membership, s is the current inventory state, p is a prototype inventory state, P is the number of prototypes, a is 1 if the object o at the terminal node of the path associated with prototype p is present and 0 otherwise, and d^{-1} is an inverse-distance or proximity function. The membership distribution implicitly specifies the agent's current goals, which will later facilitate integration with the Bayesian motion-modeling system. The final utilities are:

$$U(c) = \gamma^{e(c)} \sum V_p(c) m_p(\boldsymbol{s}), \quad c_{t+1} = \max_y U(y), \quad y \in \{x|E_{c,x} = 1\} \tag{2}$$

where $U(c)$ is the final utility of node c, γ is the discount, $e(c)$ is the number of times the player has entered cluster c, $V_p(c)$ is the original value of node c in state prototype p, and E is the edge matrix.

2.3 Bayesian Motion Modelling

It is not sufficient to simply identify the player's goals and the paths along which (s)he moved to reach them; it is also necessary to capture the actions executed by the player in pursuit of these goals. In a previous contribution [12], Thurau et al describe an approach based on Rao, Shon & Meltzoff's Bayesian inverse-model for action selection in infants and robots [14]. The choice of action at each timestep is expressed as a probability function of the subject's current position c_t, next position c_{t+1} and goal c_g:

$$P(a_t|c_t, c_{t+1}, c_g) = \frac{P(c_{t+1}|c_t, a_t)P(a_t|c_t, c_g)}{\sum_u P(c_{t+1}|c_t, a_u)P(a_u|c_t, c_g)} \tag{3}$$

This model fits into the strategic navigation system almost perfectly; the clusters c_t and c_{t+1} are chosen by examining the utility values, while the current goal state is implicitly defined by the membership distribution. In order to derive the probabilities, we read the sequence of actions taken by the player as a set of vectors $\boldsymbol{v}$ such that $\boldsymbol{v} = [\Delta\text{yaw}, \Delta\text{pitch}, \text{jump}, \text{weapon}, \text{firing}]$. We then cluster these action vectors to obtain a set of *action primitives*, each of which amalgamates a number of similar actions performed at different times into a single unit of behavior.

Several important adaptations must be made in order to use this model in the game environment. Firstly, Rao's model assumes that transitions between states are instantaneous, whereas multiple actions may be performed in QUAKE II® while moving between successive clusters; we therefore express $P(c_{t+1}|c_t, a_t)$ as a soft-distribution of all observed actions on edge $E_{ct,ct+1}$ in the topological map. Secondly, Rao assumes a single unambiguous goal, whereas we deal with multiple weighted goals in parallel. We thus perform a similar weighting of the probabilities across all active goal clusters. Finally, Rao's model assumes that each action is independent of the previous action. In QUAKE II® , however, each action is constrained by that performed on the preceding timestep; we therefore introduce an additional dependency in our calculations. The final probabilities are computed as follows:

$$\sum_g m_g P(a_t|c_t, c_{t+1}, c_g) \frac{P(a_t|a_{t-1})}{\sum_u P(a_u|a_{t-1})} \tag{4}$$

3 Believability Testing

As discussed earlier, there exists no standard method of gauging the 'believability' of game bots, nor of objectively comparing this quality in different agents; given that one of the central aims of our work lies in improving the believability of such agents, this is clearly a shortcoming which needs to be addressed. The most obvious means of determining the degree to which agents are perceived as human is to conduct a survey. This, of course, immediately raises questions of subjectivity, experimenter influence, and so on. In order to produce a credible assessment of agent believability, any proposed system must be designed with these concerns in mind. Our aims, then, are as follows: i) to construct a framework which facilitates rigorous, objective testing of the degree to which game agents are perceived as human; ii) to formulate a *believability index* expressing this 'humanness', and allowing comparisons between different agents.

The system developed to fulfil these criteria is described below. We outline the structure of the survey and its applicability to the testing of agents in general, using our own experiments to illustrate key concepts; we then describe these experiments and their results in greater detail. The test itself can be taken at `http://reynard.computing.dcu.ie/sab_tests/`

3.1 Structure of the Believability Test

To counteract any potential observer bias, the test takes the form of an anonymous online survey. Respondents are first presented with detailed instructions covering all aspects of the test. Before starting, they are further required to estimate their experience in first-person shooter games, at one of five different levels. Subjective judgements are avoided by explicitly qualifying each experience level:

1. Never played, rarely or never seen
2. Some passing familiarity (played / seen infrequently)
3. Played occasionally (monthly / every few months)
4. Played regularly (weekly)
5. Played frequently (daily)

Upon proceeding to the test itself, the respondent is present with a series of pages, each of which contains a group of video clips. Each group shows similar, but not identical, sequences of gameplay from the perspective of the in-game character. This approach was adopted due to concerns that asking respondents to view individual clips in isolation, with no basis for comparison against similar samples, would lead to a significant amount of subjectivity and guesswork. Within each group, the clips may depict any combination of human and artificial players; the respondent is required to examine the behaviour of the character in each clip, and indicate whether (s)he believes it is a human or artificial player. The clips are marked on a gradient, as follows:

1: Human, 2: Probably Human, 3: Don't Know, 4: Probably Artificial, 5: Artificial

This rating is the central conceit of the survey, and will later be used to compute the believability index. The respondent is also asked to specify how many times (s)he viewed the clip (to a maximum of 3 times), and to provide an optional comment explaining his/her choice. In cases where (s)he indicates that (s)he believes the agent to be artificial, (s)he will be further asked to rate how "humanlike" (s)he perceives its behaviour to be, on a scale of 1 to 10. This more subjective rating is not involved in the computation of the believability index, but may be used to provide additional insight into users' opinions of different agents. Having completed all required sections on each page, the user submits his/her answers and moves on to the next.

3.2 Subjectivity, Bias and Other Concerns

Aside from the observer effect, there are several areas in which the potential for subjectivity and the introduction of bias exist. Since our aim is to provide

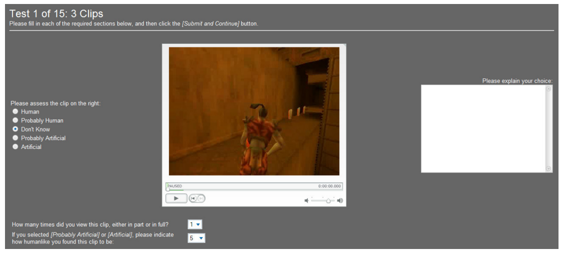

Fig. 2. Extract from the main believability test screen

an objective measure of believability, these must be eliminated or minimized. A number of these issues are discussed below.

The first obvious pitfall lies in the selection of video clips. The selector may deliberately choose certain clips in an effort to influence the respondents. To guard against this, we first ensure that the number of samples is sufficient to embody a wide variety of behaviours, and secondly, we cede control of the selection of the specific behaviours to an unbiased arbiter. In our case, we wished to compare the believability of our imitation agents against both human players and traditional rule-based bots; thus, we first ran numerous simulations with the traditional agent – over whose behavior we had no control – to generate a corpus of gameplay samples, and then proceeded to use human clips embodying similar behavior both in the believability test and to train our imitation agents.

Similarly, the order in which the videos are presented could conceivably be used to guide the respondents' answers. To prevent this, we randomize the order in which the groups of clips are displayed to each user, as well as the sequence of clips within each page; the test designer thus has no control over the order of the samples seen by the user. Additionally, the filenames under which the clips are stored are randomized, such that the respondent cannot determine the nature of each clip based on examining the webpage source (e.g. clip 1 always human, clip 2 always artificial, etc).

Another issue concerns the possibility that users will choose the 'Probably' options in a deliberate effort to artificially minimize their error and 'beat' the test, or that they will attempt to average out their answers over the course of the survey – that is, they may rate a clip as 'human' for little reason other than that they rated several previous clips as 'artificial', or vice-versa. To discourage this, we include notes on the introduction page to the effect that the test does not adhere to any averages, that the user's ratings should be based exclusively upon their perception of the character's behavior in each clip, and that the user should be as definitive as possible in their answers. A related problem is that of user fatigue; as the test progresses, the user may begin to lose interest, and will consequently invest less effort in each successive clip. We address this by including a feature enabling users to save their progress at any point, allowing them to complete the survey at their convenience.

It is also imperative to ensure that the test is focused upon the variable under investigation – namely, the believability of the agent's movement and behavior. As such, the survey must be structured so as not to present 'clues' which might influence the respondents. For instance, the tester should ensure that all clips conform to a standard presentation format, so that the respondent cannot discern between different agents based on extraneous visual cues - different players may have used different in-game character models, individual player names, etc. To this end, we run a script over the demo files to remove all such indicators and homogenize them to a common format.

In the specific case of our imitation agents, this requirement that all extraneous indicators be removed raises a conflict between two of our goals in conducting the survey. If the players in two of the three clips we use on each page begin from the same location and exhibit near-identical behavior, the respondent may conclude through pure logical deduction that (s)he is probably viewing a human and imitation agent, and consequently that the remaining clip is more likely to be a traditional artificial agent. Note that this might not necessarily be true, but even an incorrect answer based on factors other than believability will adversely affect the accuracy of the results. We circumvent this problem by training imitation agents with different (but similar) samples of human gameplay to those actually used in the test. The resulting clips are therefore comparable, but do not 'leak' any additional information; respondents must judge whether or not they are human based solely on their appearance. At the same time, however, we obviously wish to test how accurately our agents can capture the aesthetic appearances of their human exemplars. To satisfy both requirements, a small minority of imitation agents *are* trained using the same human data as presented in the survey; in the experiments described below, 2 of the imitation agents were direct clones, while the remainder were trained on different data.

3.3 Evaluation of Results

Before evaluating the results of the survey, one should ensure that there have been a substantial number of responses with a decent distribution across all experience levels; a good 'stopping criterion' is to run the test until the average experience level is at least 3 (i.e. a typical gamesplayer). Standard analyses (precision, recall, etc) can be carried out on the results; however, as mentioned earlier, we also wish to formulate a believability index which is specifically designed to express the agent's believability as a function of user experience and the certainty with which the clips were identified.

Recall that each clip is rated on a scale of 1 (definitely human) to 5 (definitely artificial). Obviously, the true value of each clip is always either 1 or 5. Thus, we can express the degree to which a clip persuaded an individual that the visualised character was human as the normalised difference between that person's rating and the value corresponding to 'artificial':

$$h_p(c_i) = \frac{|r_p(c_i) - A|}{max(h)} \tag{5}$$

where $h_p(c_i)$ is the degree to which person p regarded the clip as depicting a human, $r_p(c_i)$ is person p's rating of clip i, A is the value on the rating scale which corresponds to 'artificial', and $max(h)$ is the maximum possible difference between a clip's rating and the value of 'artificial'. In other words, $h_p(c_i)$ will be 0 if the individual identified a clip as artificial, 1 if he identified it as human, and somewhere in between if he chose one of the 'Probably' or 'Don't Know' options. We now weight this according to the individual's experience level:

$$w_p(c_i) = \frac{e_p h_p(c_i)}{avg(e)} \tag{6}$$

where e_p is the experience level of person p and $avg(e)$ is the mean experience level. Finally, we sum the weighted ratings across all clips and respondents, and take the average:

$$b = \frac{\sum_p^n \sum_i^m w_p(c_i)}{nm} \tag{7}$$

where b is the believability index, n is the number of individual respondents, and m is the number of clips. The believability index is, in essence, a weighted representation of the degree to which a given type of clip was regarded as human, in the range $(0, 1)$. In order to express the *strength* of the result and to facilitate comparison between agents evaluated in different surveys, we also compute a *confidence index* as follows:

$$c = \frac{avg(e)}{max(e)} \tag{8}$$

where $avg(e)$ is the average experience of the respondents, and $max(e)$ is the maximum experience level; the confidence index is thus conditioned upon a sufficient level of expertise among respondents. In the context of the survey, then, a 'good' result for an AI agent would involve a high value of b for both the agent and human clips, together with a confidence index of 0.6 or greater (indicating that respondents were, on average, significantly experienced).

4 Experiments

In this section, we detail an experiment carried out using the believability test in conjunction with our imitation agents. The purpose of this experiment was twofold; first, to evaluate the believability-test framework itself, and second, to examine how believable our imitation agents were in comparison with human players and traditional rule-based artificial agents.

The experiment consisted of 15 groups, with 3 clips in each; these clips were, on average, approximately 20 seconds in length. We first ran numerous simulations involving the rule-based artificial agent to derive a set of gameplay samples, and then used similar samples of human players both in the test itself and to train our imitation agents. The rule-based agent used was the QUAKE II® Gladiator bot, which was chosen due to its reputation as one of the best bots available.

Table 1. Believability/Confidence indices, Recall and Precision values. Recall values consider classification as ‘human’ to be the desired results. Precision is estimated over [human or imitation] identified as human, and rule-based agent identified as artificial.

Clip Type	Believability	Confidence	Recall (%)	Precision (%)
Human	0.69		68.08	78.39
Imitation	0.69	0.64	68.81	
Artificial	0.36		36.69	50.87

It should be noted that, since our imitation mechanism is designed to imitate strategic navigation and human motion, combat was omitted from consideration in this study. As one of our respondents commented, this filters out one variety of behavior from the agent’s repertoire, and has the effect – as with the original Turing test – of reducing the opportunities for an observer to detect artificialities. While the test can be used to accurately gauge how well our system captures human strategy and movement, a further study involving combat behaviours is essential. See Future Work for further discussion.

With the video clips in place, the URL of the survey site was distributed to the mailing lists of several colleges in Ireland and Germany. After a one-week test period, we had amassed a considerable number of responses. After discarding incomplete responses, we were left with 20 completed surveys, totalling 900 individual clip ratings; the average experience level of respondents was 3.2.

As can be seen from Tab. 1, the survey produced a very favourable impression of our imitation agents compared to the artificial agent. The believability indices for human, imitation and traditional artificial clips were 0.69, 0.69 and 0.36, respectively. In other words, the imitation agents were misidentified as human 69% of the time, while the rule-based agents were mistaken as human in only 36% of cases (weighted according to experience). Clips which actually *did* depict human players were also identified 69% the time. Essentially, it seems that respondents were generally unable to discern between the human players and our imitation agents. These results are corroborated by the recall values, which indicate that both the human and imitation clips were classified as human in approximately 68% of cases, while the rule-based agent was classified as human only 36.69% of the time. Since the human sources used to train the imitation agents were different than those human clips presented as part of the test, this implies that the results are based on the general abilities of the imitation mechanism, rather than any factors unique to the clips in question.

Further indication of the imitation agents’ effectiveness is evident in the graph of believability against experience level shown in Fig 3; as experience level rises, respondents correctly identify human clips as human more frequently, and misidentify the traditional agent as human less frequently. The identification of imitation agents as human, by contrast, closely parallels that of genuine human clips. These trends may be explained by the fact that more experienced players have a greater knowledge of characteristically human behaviours – smooth

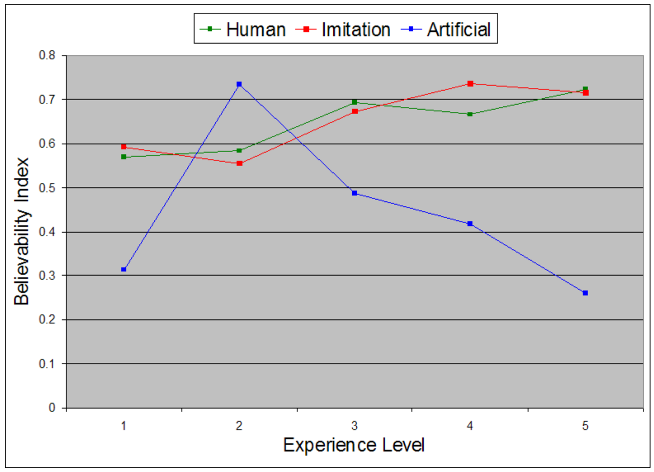

Fig. 3. Variation of believability with experience level

strafing, unnecessary jumping, pausing to examine the environment, and similar idiosyncrasies – which the traditional agent would not exhibit, but which would be captured and reproduced by the imitation bots. This interpretation is supported by many of the comments submitted by respondents, including those shown in Table 2.

Table 2. Sample comments from imitation clips misidentified as human

Experience	Comment
5	Bunny hop for no reason, also seems to be scanning for enemies
5	Fires gun for no reason , so must be human
5	Unnecessary jumping
5	Stand and wait. Ai wouldn't do this (?)
5	Human as they knew how to Rocket jump
5	The rocket jump and the short sequence of backward running at the end suggest this was human

In conclusion: while further testing (mainly of combat behaviours) is required, the results of the believability study suggest that our imitation agents exhibit far greater 'humanness' than even a well-regarded rule-based agent, and indeed are comparable to genuine human players. We consider this to be strong evidence in support of our original premise; namely, that imitation learning has the potential to produce more believable game agents than traditional AI techniques.

5 Summary & Future Work

In this paper, we proposed a formal method of quantifying the degree to which different agents are perceived as 'humanlike', in the form of a web-based survey and an objective metric based on both the respondents' level of experience and the accuracy with which the players/agents were identified. Through our experiments, we verified the effectiveness of the believability-testing system; we further showed that our imitation-learning approach produces game bots which are capable of conveying a significantly more humanlike impression than traditional agents, and are often almost indistinguishable from genuine human players.

Clearly, the next stage in our work must concentrate on imitating combat behaviours, and integrating them into the existing imitation mechanism. Beyond this, tests based on the third metric described in the introduction will also be conducted – that is, in-game performance-based evaluation of the imitation bots, in direct competition with human players and other artificial agents.

References

1. Schaal, S.: Is Imitation Learning the Route to Humanoid Robots? Trends in Cognitive Sciences **3**(6) (1999) 233–242
2. Hayes, G., Demiris, J.: A Robot Controller Using Learning by Imitation. In: Proc. of the 2nd Int. Symposium on Intelligent Robotic Systems. (1994) 198–204
3. Schaal, S.: Movement Planning and Imitation by Shaping Nonlinear Attractors. In: 12th Yale Workshop On Adaptive And Learning Systems. (2003)
4. Fod, A., Matarić, M., Jenkins, O.: Automated Derivation of Primitives for Movement Classification. Autonomous Robots **12**(1) (2002) 39–54
5. Laird, J.E., v. Lent, M.: Interactice Computer Games: Human-Level AI's Killer Application. In: Proc. AAAI. (2000) 1171–1178
6. Fairclough, C., Fagan, M., MacNamee, B., Cunningham, P.: Research Directions for AI in Computer Games. Technical report, Trinity College Dublin (2001)
7. Livingstone, D., McGlinchey, S.: What Believability Testing Can Tell Us. In: In Proc. Int. Conf. on Computer Games: AI, Design and Education. (2004)
8. Livingstone, D.: Turing's Test and Believable AI in Games. CiE **4**(1) (2006) 6
9. J.E. Laird, J.D.: Creating Human-like Synthetic Characters with Multiple Skill-Levels: A Case Study Using the Soar Quakebot. In: Proc AAAI. (2000)
10. Gorman, B., Fredriksson, M., Humphrys, M.: QASE – An Integrated API for Imitation and General AI Research in Commercial Computer Games. In: Proc. CGAMES Int. Conf. Computer Games. (2005) 207–214
11. Gorman, B., Humphrys, M.: Towards Integration of Strategic Planning and Motion Modelling in Interactive Computer Games. In: Proc. Int. Conf. Computer Game Design & Technology. (2005) 92–99
12. Thurau, C., Paczian, T., Bauckhage, C.: Is Bayesian Imitation Learning the Route to Believable Gamebots? In: Proc. GAME-ON North America. (2005) 3–9
13. Laird, J.E.: Using a Game to Develop Advanced AI. IEEE Computer (2001) 70–75
14. Rao, R., Shon, A., Meltzoff, A.: A Bayesian Model of Imitation in Infants and Robots. In Dautenhahn, K., Nehaniv, C., eds.: Imitation and Social Learning in Robots, Humans, and Animals: Behavioural, Social and Communicative Dimensions. Cambridge University Press, (2004)