On the simulation of interactive non-verbal behaviour in virtual humans

John Shearer[1] and Patrick Olivier[1] and Marco De Boni[2]

Abstract Development of virtual humans has focused mainly in two broad areas – conversational agents and computer game characters. Computer game characters have traditionally been action-oriented – focused on the game-play – and conversational agents have been focused on sensible/intelligent conversation. While virtual humans have incorporated some form of non-verbal behaviour, this has been quite limited and more importantly not connected or connected very loosely with the behaviour of a real human interacting with the virtual human – due to a lack of sensor data and no system to respond to that data. The interactional aspect of non-verbal behaviour is highly important in human-human interactions and previous research has demonstrated that people treat media (and therefore virtual humans) as real people, and so interactive non-verbal behaviour is also important in the development of virtual humans. This paper presents the challenges in creating virtual humans that are non-verbally interactive and drawing corollaries with the development history of control systems in robotics presents some approaches to solving these challenges – specifically using behaviour based systems - and shows how an order of magnitude increase in response time of virtual humans in conversation can be obtained and that the development of rapidly responding non-verbal behaviours can start with just a few behaviours with more behaviours added without difficulty later in development.

1 INTRODUCTION

Interactive non-verbal behaviour is important in human-human interaction, but has to date been given very limited attention by the virtual human AI community. AI in games has been focused more on game play with attention only recently towards non-verbal communication in games such as Half-Life 2 [1]. Previous games had non-verbal communication limited to cut-scenes. AI researchers have been focused on conversation for a long time, but mainly under a natural language processing paradigm – that is trying to understand spoken (or more often textual) language and respond appropriately[2]. More recently virtual humans capable for full body expression have been developed and these have proved engaging[3, 4, 5]. Their limitation has been that similar to simpler text-based or speech-based systems their only input has been typed or spoken speech. The non-verbal behaviour has therefore only been based on the textual input and output, ignoring the important behaviour in the non-verbal modality (though [6] and similar research attempts with some success to predict non-verbal behaviour based on the speech modality only as [7, 8, 9] show significant redundancy between the modalities). Non-verbal behaviour, especially gesture, has been given attention under a computer control paradigm [10] and also inform interactive system as a whole [11], but little attention has been given to the actual development of virtual humans that utilise non-verbal behaviour as both input and output, especially in a fast control loop. The notable exception to this is [12] who use head nod detection for conversational feedback – to inform the flow of conversation..

The introduction of more complex data streams to virtual humans introduces difficulties with the analysis of this data and also with determining appropriate behaviour based on this input data. Present AI systems in virtual humans are either very simple rule based systems, such as those in computer games or imitation agents [13], or highly complex natural language processing (NLP) systems that attempt to fully understand the context of spoken or more usually typed language and search for appropriate actions. Fully modelling the world and searching for appropriate actions has been possible due the limited form of data input. The additional complexity and unpredictability of non-verbal behaviour input introduces similar problems to AI systems for virtual humans that were approached in the 1980's for AI systems for robot control. The use of a full sense-process-act cycle for the AI systems was too complex and more importantly too *slow* for real-time systems (such as robotics, or interactive virtual humans). All virtual systems at present have a response time of at least half a second, and many much more (text systems usually only respond when new text is input).

In comparison with robotics AI history, the real-time behaviour of virtual humans is still in the first stage of development (sense-plan-act - which worked for robotics in simulated or highly restricted environments, and is still appropriate in many circumstances). In order for virtual humans to be interactive in real-environments their behaviour response time need to be reduced by an order of magnitude – towards that of humans in normal conversation. That is, they need to response immediately to a users behaviour, which is not to say that their full response must be immediate, but that there must be *some* immediate response. We propose drawing on the further developmental stages of real-time robotics AI systems to provide inspiration for virtual human AI systems – specifically subsumption architecture[14] and behaviour based systems, moving towards more hybrid systems[15] drawing on the strengths of present virtual human AI systems with the addition of simpler fast response behaviours. These stages of robot

[1] Newcastle University, Culture Lab, Kings Walk, Newcastle University, NE1 7RU, UK, email: john.shearer@ncl.ac.uk; p.l.olivier@ncl.ac.uk
[2] Unilever Corporate Research, Unilever R&D Colworth, Sharnbrook, Bedford, MK44 1LQ, UK, email: marco.de-boni@unilever.com

AI systems made it possible, in addition to increased response times, to build up robot behaviours step-by-step with increased reliability and robustness using less computing power than previously thought possible. We believe that it is possible to build up a fully interactive virtual human using a hybrid approach of behaviour based systems and the more traditional virtual human techniques, but at this state the focus in on developing early prototypes that interact in simple ways before moving towards more complex systems.

The next section provides more detail and history of the development of AI control systems in robotics along with the advantages and disadvantages of these approaches. Section 3 shows how these developments can be applied in virtual humans and discusses the importance of conversational state in interactions and that the relative context-freeness (from the specific high-level conversation meaning) enables that behaviours can be modulated by the conversational state without awareness of that high-level context. We then provide some details on the present state of development our behaviour based virtual human system and discuss how it is possible to initially build as system with just a few behaviours, with further behaviours being able to be added at a later a date without difficulty. Finally moving on to some approaches to evaluating these virtual humans, both in their entirety and piecewise (i.e. evaluating *which* behaviours are important).

2 DEVELOPMENTAL HISTORY OF AI CONTROL SYSTEMS IN ROBOTICS

Norbert Weiner in the late 1940s developed the field of cybernetics – the "marriage of control theory, information science, and biology that seeks to explain the common principles of control and communication in both animals and machines"[16] – which affirmed the notion of situatedness – the strong two-way coupling between an organism and its environment[16]. It is this strong two-way coupling that seems to be missing from present state-of-the-art virtual humans. There is, of course, two-way coupling in all virtual humans. The difficulty lies with the limited strength of that coupling. The focus of this paper is on the limitation of the coupling in terms of the limited sensory input and the limited response speed – both contributing to the limited strength of the coupling. We should note at this point that there are other factors that reduce the strength of the coupling as compared with that of real human-human interactions, such as the lack of physicality, realism, etc in virtual humans.

Following on from Weiner's work W. Grey Walter designed and constructed some of the earliest robots using simple sensors and actuators (and entirely analogue computing), with strong coupling between those sensors and actuators [17]. These simple machines, consisting merely of two sensors (a photocell and a bump sensor), two actuators (motors), and two "nerve cells" (vacuum tubes) were capable of surprisingly complex behaviour – seeking light, heading towards a weak light, back away from a bright light, etc. For whatever reasons this work was not strongly continued until revived almost 30 years later by Braitenberg [18] as a series of thought experiments, which were eventually transformed into true robots. MIT's Media Lab built twelve such robots and demonstrated a large variety of simple behaviours, including a timed shadow seeker, an indecisive shadow-edge finder, a paranoid shadow-fearing robot and a driven light seeker [19].

It is generally held that the start of artificial intelligence (AI) as a separate field was associated with a summer research conference held at Dartmouth University in 1955, with the original proposal indicating that an intelligent machine "would tend to build up within itself an abstract model of the environment in which it is placed. If it were given a problem it could first explore solutions within the internal abstract model of the environment and then attempt external experiments" [20]. From this point onwards the dominant approach in robotics and AI research for the next three decades was this representational knowledge and deliberative reasoning approach - representing hierarchical structure by abstraction; and using "strong" knowledge employing explicit symbolic representational assertions about the world.

In [21], Brooks claimed that "planning is just a way of avoiding figuring out what to do next", and while that is perhaps a little extreme, it does embody the idea of behaviour based systems and exemplifies the reaction against the traditions of classical AI. At this point also, advances in robotic hardware made it feasible to test the behaviour based approaches in real robots. The area of distributed artificial intelligence (DAI) developed at or around the same time as behaviour based systems in robotics. The idea that multiple competing or cooperating processes (or demons/daemons, or agents) could generate coherent behaviour [22, 23, 24], and Arkin states "individual behaviours can often be viewed as independent agents in behaviour based robotics, relating it closely to DAI" [25].

Approaches and techniques for robotics control can be depicted in on a spectrum from deliberative system to reactive systems as in Figure 0 ([25], page 20). As discussed previously, other than in computer games the focus for humans has been towards the deliberative end of this spectrum – developing virtual humans with well developed high-level level intelligence abilities, but as shown in the diagram these more cognitive process have a slower response time. As each person knows from their own normal lives, interactions with other people are made up of a whole set of different responses that sit along the deliberative-reactive spectrum, and all these varied responses are important for a smooth and useful interaction, not just the high-level responses. Therefore, a virtual human (like most present day ones) that only exhibits high-level intelligence is missing out on important low-level intelligence, which is also important. The relative importance of the levels of intelligence is clearly variable and is not under discussion here, but it is clear from a long history of work is psychology that these lower-level intelligence responses, such as eye-contact, intonation, gesture, back-channel speech, are highly important in human interactions, and therefore also in human-virtual human interactions [26, 27, 28, 29, 30]. The structure of human (and other animal) brains reflects this continuum from simple to complex behaviours and while the physical separation of different parts of the brain for different behaviours was part of the inspiration for

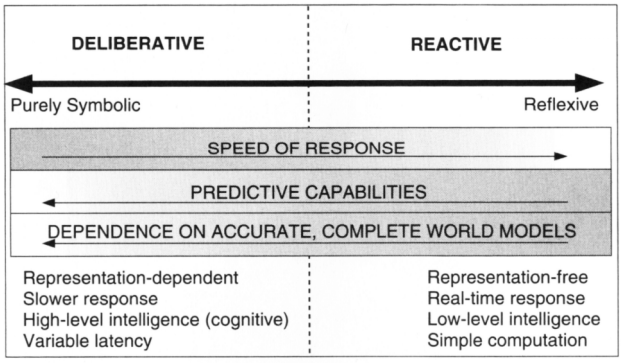

Figure 0 - Robot control systems spectrum

behaviour based systems, behaviour based system do not claim to be a replication or model of the human (or any animal) brain, merely drawing on them for ideas.

Robots (or virtual humans) utilising deliberative reasoning require relatively complete knowledge about the world and tend to struggle in more dynamic worlds where data that the reasoning processes uses may be inaccurate or have changed since last reading. More importantly, the deliberative reasoning process is frequently slow. Behaviour based systems or reactive systems were developed to attempt to solve some of the apparent drawbacks of deliberative systems – namely a lack of responsiveness in unstructured and uncertain environments.

A reactive control or a behaviour is a simply a tight coupling between perception and action to produce timely responses in dynamic and unstructured worlds. A behaviour based system is a collection of behaviours (perception-actions) pairs that cooperate/compete to produce more global behaviour. The obvious difficulty with having multiple behaviours is how to choose which behaviours should take control in times of conflict. The approach usually used in behaviour based systems in simply a priority system where higher priority behaviours win out over lower priority behaviours. The idea of one behaviour winning out over another (lower priority] behaviour also applies, in addition to behavioural outputs, to behavioural inputs. That is, rather than there existing a separate conflict "resolver" choosing between the outputs of behaviours A (high priority) and B (low priority), view behaviour A as inhibiting, or replacing the outputs of B. It is then, a relatively small leap to imagine that behaviour A could also inhibit or replace the *inputs* of B. This is the idea of Brooks' "subsumption architecture" [14].

Within the field of robotics behaviour based systems saw significant success before running into the problem that almost inevitably, without any high-level or abstract representations the systems were incapable of the more complex behaviours that we wanted. The obvious next step was a hybrid between the two where behaviour based systems provide the fast, reactive control, while the deliberative systems provide the slower higher level cognitive control[15]. And it would be perhaps fair to say that many people would not view a robot or a virtual human with *only either* fast reactions *or* high level cognitive behaviours as intelligent – it would be both.

3 BEHAVIOUR BASED ARCHITECTURES FOR VIRTUAL HUMANS AND CONTEXT-FREE BEHAVIOURS

A behaviour based system consists of a set of behaviours, some of which can subsume (override or replace) the inputs and/or outputs of others (inhibition is simple overriding with nothing). We can view even slow high-level cognitive processes as behaviours, and therefore present deliberative virtual human control systems are simple behaviour based systems with one (or a few) complex behaviours, and furthermore a hybrid system is also just a behaviour based system. Behaviour based systems as applied to robots usually apply the behaviours directly to drive systems (motors, etc.). While this is possible in virtual humans (to control joint angles, muscle forces, etc.), it is also possible for a behaviour based system to control at a higher level – i.e. control the various animations that a virtual human may already have. This is the main adaptation needed to apply behaviour based systems to virtual humans.

Within human interactions the lower-level behaviours are predominantly unaware of the deeper meanings in an interaction and are consistent across different interactional contexts. In other words whether an interaction involves talking about the weather; discussing the latest cricket result; who ate all the pies; or solving world hunger, the majority of human interactional behaviours are still present and the same – i.e. people still look at each other (enough, but not too much); they still nod in agreement (in western cultures); and still give back-channel speech encouragers, etc. Of course, not all these behaviours are present all the time and are sometimes affected by high-level context, for the most part they are not. That said; these behaviours are influenced by the conversational state. This is the state of conversation from the simple state of whose turn it is to speak, to the deeper levels of state such as "Bill is speaking, but Ted is trying to break into the conversation". These conversational states influence the various behaviours that are active (or their form). For example, as Ted is trying to break into the conversation, Bill will have increased behaviour(s) that try to hold the turn. In other conversational states Bill will have other behaviours enabled and disabled.

As one might expect the conversational state is again just a more complex behaviour or set of behaviours, with transitions between states caused by sensory input. So, this fits nicely into the whole behaviour based model – the conversation state behaviour modulates (or subsumes) some of the lower level behaviours.

Before moving on to some implementation details of behaviour based systems with virtual humans we should note that the idea of having rapidly interactive virtual humans has been worked on in the field of animation, especially by Perlin [31]. The main limitation of this work is that it was not grounded in behaviours and behavioural responses that real people use and that it did not investigate the scalability of the solutions (which behaviour based robotics has). It was found that character that react quickly and variedly to people were engaging and appeared to portray personality.

4 DEVELOPMENTS WITH BEHAVIOUR BASED VIRTUAL HUMANS

In practice when connected together a set of behaviours create a directed graph between input, output, and processing elements. The ideas of subsumption (one element overriding another's inputs and/or outputs) can be implemented by redirecting the edges within that graph. The idea of a graph of processing elements has been implemented in a variety of multimedia processing frameworks. Both DirectShow [32] on Windows and GStreamer [33] on Linux and Windows connected elements into pipelines or directed graphs. Additionally, the EyesWeb open platform [34] utilises a directed graph approach to supporting multimodal expressive interfaces and multimedia interactive systems and uses a visual programming paradigm whereby elements can be placed and connected together in a GUI. This visual programming paradigm is also present in both DirectShow and GStreamer. The advantage of EyesWeb is that it includes significant elements for performing both complex vision (OpenCV [35]) and audio processing, which is needed in order for a virtual character to respond to real-world sensory data.

For our early investigations into using behaviour based systems to control virtual humans, our virtual human [36] was adapted to be accessible from EyesWeb and we then designed simple vision and audio processing graphs (or pipelines) to control the character. We found that it is easy to create simple reactive behaviours, and the response time of the system is fast as it is only limited by the processing speed and the latencies of the hardware – there is no high level processing occurring at this point. It is no surprise that the main difficulties lie with the vision and audio processing – i.e. managing to detect the right things, but it is easy to add significant sensory capability in this system. The actual behavioural parts are straightforward, and it is simply a matter of moving some of the connections to subsume lower level behaviours. The follow on stage involves adding a larger set of detectable human behaviours and responses behaviours, followed by the modulation of these behaviours by the conversation state behaviour. We will also be using additional sources of interactional data, such as eye tracking. Further work will be reported at a later date, but behaviours are independent apart from their inputs and outputs being subsumable. Therefore adding additional behaviours does not invalid the previous ones – they can just be added in, subsuming other behaviours when needed.

5 EVALUATING BEHAVIOUR IN BEHAVIOUR BASED VIRTUAL HUMANS

General evaluation of virtual humans has been relatively limited to date [37, 38] and is dependent upon definitions of what metrics make a "good" virtual human and this varies with context. Within any specific domain metrics can be created to measure the important aspect within that domain, for example, how much people like the virtual human. But, the focus in this paper is not on evaluating virtual humans generally, but on how to evaluate a) whether a virtual human with these additional simple, fast-acting behaviours is better, and b) which of those behaviours help the most. Both these evaluations could be run together. Assuming one had an appropriate metric, the virtual human could be tested with a variety of combinations of behaviours on and off - including all behaviours expect the high-level cognitive behaviours off (i.e. a virtual human like present ones), vice versa (how good is a virtual human with *only* simple behaviours?), and any other combinations. Statistical analysis will allow the determination of the quality contributions of the individual behaviours. The knowledge of which behaviours are important will be useful not only to inform which behaviours to focus on in terms of development or in more limited systems, but also useful to inform (or be a test bed for) areas such as psychology which behaviours are especially important in human-human interactions. This could be especially useful for people suffering from various forms of autism – both to inform which behaviours they could focus on, but also to provide a transparent systems where they could see how and why it responds as it does. Finally, we haven't discussed or tried how these virtual human would respond to each others more varied set of behaviours. This is something that could be highly interesting to investigate in the future, and how interactions that are interesting or realistic to a third party observe could be based on *only* simple behaviours.

REFERENCES

[1] Valve Corporation, *Half-Life 2*, Valve Corporation, Bellevue, Washington, 2004.

[2] J. Weizenbaum, *ELIZA - a computer program for the study of natural language communication between man and machine*, Communications of the ACM, 9 (1966), pp. 36-45.

[3] J. Cassell, H. H. Vilhjálmsson and T. Bickmore, *BEAT: the Behavior Expression Animation Toolkit*, *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM Press, 2001.

[4] P. Tepper, S. Kopp and J. Cassell, *Content in Context: Generating Language and Iconic Gesture without a Gestionary*, *Workshop on Balanced Perception and Action in ECAs at AAMAS*, 2004.

[5] J.-C. Martin, R. Niewiadomski, L. Devillers, S. Buisine and C. Pelachaud, *Multimodal complex emotions: Gesture expressivity and blended facial expressions*, International Journal of Humanoid Robotics, Special Edition "Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids" (2006).

[6] H. Yan, *Paired Speech and Gesture Generation in Embodied Conversational Agents*, *Media Arts and Sciences, School of Architecture and planning*, Massachusetts Institute of Technology, Cambridge, MA, USA, 2000.

[7] A. Kendon, *Gesticulation and speech: Two aspects of the process of the utterance*, in M. R. Key, ed., *The Relation*

Between Verbal and Non-Verbal Communication, Mouton, The Hague, The Netherlands, 1980.

[8] D. McNeill, *Gesture and thought*, University of Chicago Press, Chicago, IL, 2005.

[9] Y. Xiong and F. Quek, *Gestural Hand Motion Oscillation and Symmetries for Multimodal Discourse: Detection and Analysis*, *Computer Vision and Pattern Recognition for Human Computer Interaction (CVPRHCI)*, Monona Terrace Convention Center, Madison, Wisconsin, USA, 2003.

[10] M. M. Cerney, *From gesture recognition to functional motion analysis: Quantitative techniques for the application and evaluation of human motion*, Iowa State University, Ames, 2005.

[11] H. Gunes, M. Piccardi and T. Jan, *Face and body gesture recognition for a vision-based multimodal analyzer*, *Pan-Sydney area workshop on Visual information processing: CRPIT '36*, Australian Computer Society, Inc., 2004.

[12] C. Sidner, C. Lee, L.-P. Morency and C. Forlines, *The Effect of Head-Nod Recognition in Human-Robot Conversation*, *Human-Robot Interaction*, Salt Lake City, Utah, USA, 2006.

[13] S. Kopp, T. Sowa and I. Wachsmuth, *Imitation Games with an Artificial Agent: From Mimicking to Understanding Shape-Related Iconic Gestures*, in V. Camurri, ed., *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 2004.

[14] R. A. Brooks, *A robust layered control system for a mobile robot*, Robotics and Automation, IEEE Journal of, 2 (1986), pp. 14-23.

[15] E. Gat, *On Three-layer architectures* in D. Kortenkamp, R. P. Bonnasso and R. Murphy, eds., *Artificial intelligence and mobile robots: case studies of successful robot systems* MIT Press, 1998.

[16] N. Wiener, *Cybernetics; or, Control and communication in the animal and the machine*, Wiley; Hermann et Cie, New York, Paris,, 1948.

[17] W. G. Walter, *The Living Brain*, W W Norton & Co, 1953.

[18] V. Braitenberg, *Vehicles, experiments in synthetic psychology*, MIT Press, Cambridge, Mass., 1984.

[19] D. W. Hogg, F. Martin, M. Resnick and Massachusetts Institute of Technology. Epistemology & Learning Research Group., *Braitenberg creatures*, Epistemology and Learning Group MIT Media Laboratory, Cambridge, MA, 1991.

[20] J. McCarthy, L. Minsky, N. Rochester and C. E. Shannon, *A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE*, 1955.

[21] Brooks, *Planning is just a way of avoiding figuring out what to do next*, Massachusetts Institute of Technology, 1987.

[22] G. Selfridge and U. Neisser, *Pattern Recognition by Machine*, Scientific American, 203 (1960), pp. 60-68.

[23] L. Erman, F. Hayes-Roth, V. Lesser and D. Reddy, *The Hearsay II Speech Understanding system: Integrating Knowledge to Resolve Uncertainty*, Computing Surveys, 12 (1980), pp. 213-53.

[24] M. L. Minsky, *The society of mind*, Simon and Schuster, New York, 1986.

[25] R. C. Arkin, *Behavior-based robotics*, Mit Press, Cambridge Mass, 1998.

[26] S. Duncan and D. W. Fiske, *Face-to-face interaction : research, methods, and theory*, L. Erlbaum Associates, Hillsdale, N.J. , 1977.

[27] M. L. Knapp and J. A. Daly, *Handbook of interpersonal communication*, SAGE Publications, Thousand Oaks, CA, 2002.

[28] L. Cerrato and M. Skhiri, *Analysis and measurement of communicative gestures in human dialogues*, *AVSP*, St. Jorioz, France, 2003.

[29] D. Efron, *Gesture and environment*, King's Crown Press, New York, 1941.

[30] D. McNeill, *Hand and mind : what gestures reveal about thought*, University of Chicago Press, 1992.

[31] K. Perlin, *Real Time Responsive Animation with Personality*, IEEE Transactions on Visualization and Computer Graphics, 1 (1995), pp. 5-15.

[32] Microsoft Corporation, *Microsoft DirectShow 9.0*, 2007 http://msdn2.microsoft.com/engb/library/ms783323.aspx.

[33] *GStreamer*, (2007), http://gstreamer.freedesktop.org/.

[34] A. Camurri, P. Coletta, A. Massari, B. Mazzarino, M. Peri, M. Ricchetti, A. Ricci and G. Volpe, *Toward real-time multimodal processing: EyesWeb 4.0*, *AISB 2004 Convention: Motion, Emotion and Cognition*, Leeds, UK, 2004.

[35] Intel Corporation, *OpenCV*, http://sourceforge.net, 2005.

[36] I. Haptek (2007), http://www.haptek.com/corporate/.

[37] Z. Ruttkay and C. Pelachaud, eds., *From Brows to Trust*, Kluwer Academic Publishers, Dordrecht, 2004.

[38] D. M. Dehn and S. v. Mulken, *The impact of animated interface agents: a review of empirical research*, International Journal of Human-Computer Studies, 52 (2000), pp. 1-22.