

Levels of Detail for Crowds and Groups.

C. O’Sullivan[†], J. Cassell[§], H. Vilhjálms[§], J. Dingliana[†], S. Dobbyn[†], B. McNamee[†], C. Peters[†] and T. Giang[†]

[†]Trinity College Dublin, [§]MIT Media Lab.

Abstract

Work on levels of detail for human simulation has occurred mainly on a geometrical level, either by reducing the numbers of polygons representing a virtual human, or replacing them with a two-dimensional imposter. Approaches that reduce the complexity of motions generated have also been proposed. In this paper, we describe ongoing development of a framework for Adaptive Level Of Detail for Human Animation (ALPHA), which incorporates levels of detail for not only geometry and motion, but also includes a complexity gradient for natural behaviour, both conversational and social.

Keywords: crowd animation, group animation, real-time animation, human simulation, collision handling

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Three-Dimensional Graphics and Realism]: Animation

1. Introduction

Crowd and group simulations are becoming increasingly important in the entertainment industry. In movies, they can be used to simulate the presence of real humans. For example, in the movie *Titanic*, extensive use was made of virtual people, whose movements were generated from a library of pre-captured motions. Such technology can be used in situations where it is dangerous for real people to perform the actions, such as falling over 50 feet off a ship, or to reduce the complexity and expense of handling large numbers of human extras. In animated movies, crowd simulation really comes into its own, such as in the stampede scene in Disney’s *The Lion King* and the colonies of ants in Dreamworks’ *Antz*. Crowd and group simulations are becoming more popular in simulated online communities and interactive applications such as games and Virtual Reality.

Recent research into crowd simulation has to a large extent been inspired by the flocking work of Craig Reynolds²⁴. He revolutionised the animation of flocks of animals, in his case birds (or “Boids”), by adapting some ideas from particle systems, where each individual bird is a particle. More recently, groups of creatures travelling in close proximity have been simulated, whose movements are controlled by dynamical laws⁴. A key element of this type of animation is collision avoidance. Significant research has been conducted into the area of crowd simulation, an example being the work of



Figure 1: Natural conversation in a group.

Musse *et al.*^{19,20}. In their ViCrowd system, they can create a virtual crowd, where the individuals have variable levels of autonomy i.e. scripted, rule-based, or guided interactively by the user. They have demonstrated the emergent behaviour of a crowd attending, for example, a political demonstration or a football match. Nevertheless, the realistic simulation of crowds and groups of humans remains a challenge, due to the ability of humans to detect even slightly unnatural human

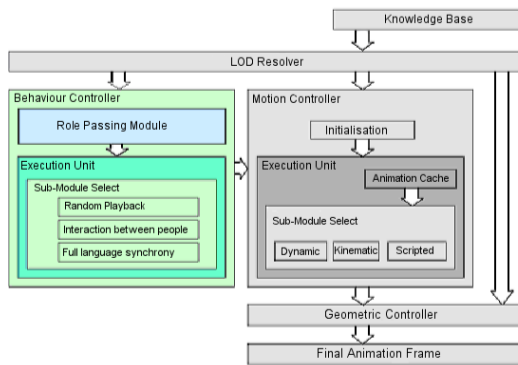


Figure 2: Extended ALOHA framework with LODs for geometry, motion and behaviour.

behaviour. In particular, realistic gesture and interactions between the individuals in the crowd are important cues for realism. Individuals in the crowd should look as if they are conversing or communicating with each other in both verbal and non-verbal ways (see Figure 1).

The crowd effects in the movie *Antz* were impressive. An extensive library of motion-tracked movements was used to assign random movement to the individuals in the crowd. Such a method looks good from a distance, but up closer it is obvious that the characters are behaving in a cyclical manner and not interacting naturally with each other. In the movie, the main characters were animated manually by human animators, which would just not be possible in any kind of real-time, interactive application. Such random behaviour may be retained as a lower level of detail, and increasingly realistic (and hence computationally expensive) techniques may be used for more important characters, *e.g.* those closer to the viewer, or characters that are more significant to the plot. Some work has been done by Aubel *et al.*¹ and Tecchia *et al.*^{26,25} on methods to reduce the rendering requirements of crowd simulations, but the approach taken there was to replace a full geometrical model with a lower resolution one, such as a two-dimensional imposter. Giang *et al.* proposed a framework for Adaptive Level Of detail for Human Animation (ALOHA)¹⁰, incorporating levels of detail for geometry, motion, and behaviour. In this paper, we describe the ongoing development of this framework, to which we have added a new behaviour controller (see Figure 2).

In Section 2 we describe the technique that we use for varying the geometrical levels of detail of our humans for rendering purposes. In Section 3 we describe how we are adding simulation levels of detail to the framework, both for generating the motions of the humans and for collision handling. Before concluding with plans for future work, Section 4 describes the behavioural levels of detail in the system, in-

roducing a complexity gradient for both conversational and more general behaviour.

2. Geometry

The requirement in interactive systems for real-time frame rates means that a limited number of polygons can be displayed by the graphics engine in each frame of a simulation. Therefore, meshes with a high polygon count often have to be simplified in order to achieve acceptable display rates. This can be achieved in several ways: a representation of the object at several levels of detail with a fixed polygon count can be generated, although switching between such levels of detail can cause a perceivable pop in an animation, or multi-resolution or progressive meshes that can be refined at run-time can be used. Alternatively, image-based imposters can be used by replacing the entire geometry of the character with images that have been taken from multiple angles²⁶. While the latter technique is excellent for visualising large crowds in real-time, the representations are not detailed enough for closer viewing, being simply flat textures. In addition, severe popping effects occur when switching from the imposter to a full geometrical model.

Subdivision surfaces offer a good solution for increasing and reducing the detail of characters in games, as demonstrated recently by Leeson¹⁶. High end rendering packages offer good support for these surfaces, and companies like Pixar have used them to produce very effective results⁷. Many of the problems associated with other curved surface representations such as NURBS can be avoided by using subdivision schemes and they behave in a way similar to polygonal meshes. Character skins can be used almost directly with some subdivision schemes while others require modification in order to get a good representation of the original mesh. We are using subdivision surfaces as a means of improving the appearance of our virtual humans, along with some acceleration methods based on culling and preprocessing. We have found that one of the advantages of subdivision surfaces is that only a low resolution mesh is necessary as the starting point - see Figure 3.

Subdivision schemes use a mask to define a set of vertices and corresponding weights, which are used to create new vertices or modify existing ones. Different masks are used for vertices on a boundary than the rest of the surface, because on a boundary edge some neighbouring vertices are absent. Other masks are also used to generate creases in a surface, making them capable of having sharp features. The masks are applied to each vertex in the mesh to produce a new mesh (and can also be applied to the texture coordinates). After successive applications of the mask the mesh converges to a surface. Figure 4 shows examples of some subdivision schemes that can be used: Linear, Loop¹⁷ and Butterfly^{9,27}. In this way, detailed meshes can be generated from a small number of preliminary vertices - see Figure 5.

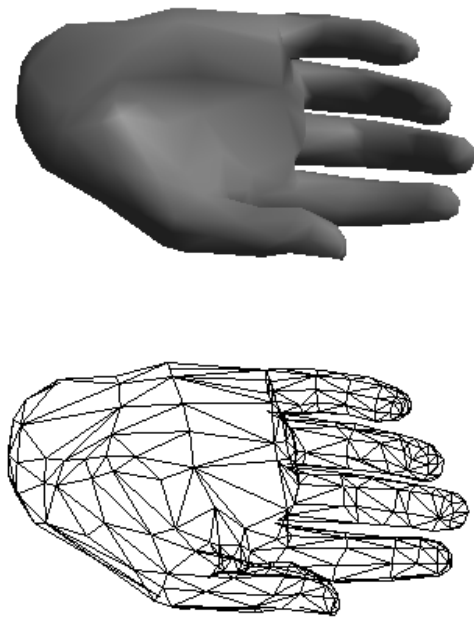


Figure 3: Subdivision schemes begin with a low-resolution model.

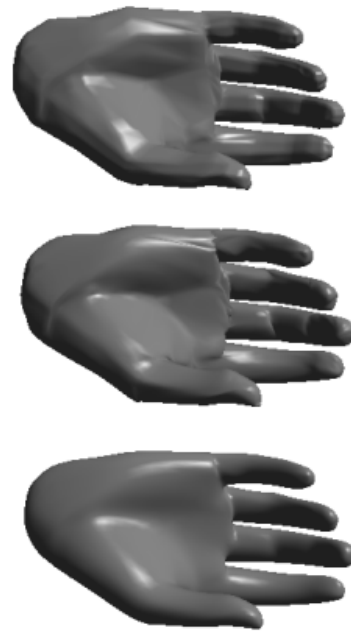


Figure 5: High resolution models created by (from top to bottom) the linear, butterfly and loop subdivision schemes.

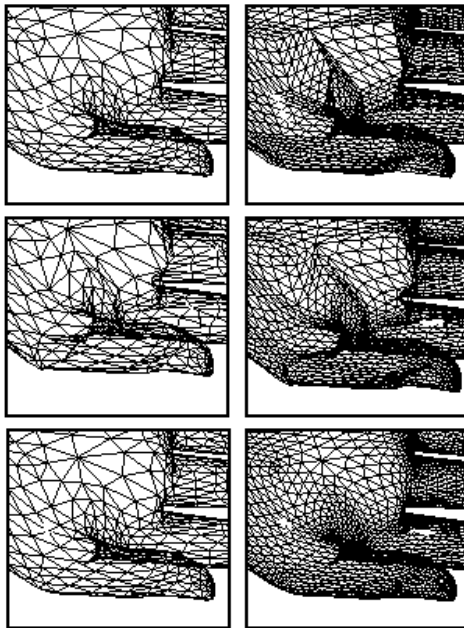


Figure 4: From left to right: two iterations of the linear, butterfly and loop subdivision techniques

3. Motion

At the lowest level of detail in ALOHA, key-framed animations, created in Character Studio or motion-captured, are exported from 3D Studio Max and used to animate the Virtual Humans. When the characters are far from the viewer, these animations are chosen at random and changed at different intervals, in order to give an impression of varied activity to the crowd. When the viewer focusses on these characters (e.g. by moving closer or starting an interaction), actions that are more meaningful are then chosen. This means applying more sophisticated behaviour, as described in Section 4, but also generating more realistic motions that do not cycle and become more accurate as the character increases in importance.

One approach we are taking to making motion look more realistic is by using a real-time reaching and grasping system. The virtual human is endowed with a memory model and has the ability to sense their environment using a virtual vision sensor. Virtual humans are not automatically aware of the exact characteristics of all objects in their environment or visual field, and must pay attention to them in order to be able to sample their attributes at a lower granularity. Given a command (e.g. “pick up the bottle”), the virtual human will become attentive towards the object and will not only generate a goal-directed arm motion towards it, but will also remember the position and other perceived attributes

of the object. Future requests regarding the object can then be dealt with using the memorised state of the object; they will be able to make realistic reaching movements towards the remembered object location without having to look at it. The generation of these arm motions is based on results from neurophysiology. See Peters and O'Sullivan (2002)²³ for further details.

Another reason why human crowd simulations tend to lack realism is because of the flocking approach often adopted. Most crowd simulations implement only collision avoidance and use coarse approximations, such as bounding boxes or spheres, to detect possible future collisions. Such approaches tend to focus on the macroscopic motions of individuals and the crowd as a whole and are suitable models for simulating distant views of crowds when they are in motion. However when a closer, more detailed view is required, this can lead to simulations where the individual elements in the crowd are too sparse and therefore unconvincing. People in a large crowd, perhaps attending a concert or sporting event, may appear to move in a flocking manner at times. However, when they are observed more closely or all group together into an enclosed area such as a bar, more accurate collision handling is needed. Real people bump up against each other, or are squashed together in an enclosed area. Even if they are not too tightly packed, people who know each other will be walking along chatting, holding hands, or occasionally touching. To model such interactions, we need to incorporate proper collision detection and appropriate responses, based on behavioural rules, in order to achieve a more chaotic, less military look to the crowds.

At the other extreme of the spectrum, approaches for highly detailed collision detection with virtual humans have been implemented, performing polygon level checks between humans and objects in the scene³. Although this might be appropriate for offline simulations and for smaller, less populated scenes requiring high detail collision information *e.g.* for cloth simulation, this obviously becomes computationally infeasible for a large scene with many virtual humans interacting and being simulated accurately at the same time. A real-time adaptive approach is most suitable for this situation, and techniques for adaptive collision detection and contact modelling, as described by Dingliana and O'Sullivan⁸, can be adapted to deal with this situation.

A hierarchy of rigid bounding volumes is used for collision detection in our system (see Figure 6). If the characters are modelled based on hierarchical transforms then these transforms can easily be used to update the positions of nodes in the collision detection hierarchy. A hierarchy of Sphere Swept Volumes (SSV's)¹⁵ is ideal for hierarchically modelled virtual humans as the volume nodes are a good fit for most of the nodes in the transformation hierarchy. Line swept spheres (LSS's) are the volumes generated by sweeping a sphere across a line segment. A hierarchy of SSV's is an efficient way to model characters for colli-

sion detection as a combination of simple spheres (referred to as Point Swept Spheres or PSS) and LSS nodes provides a good balance between computational efficiency and tight-boundedness in the case of virtual humans. Such bounding volume hierarchies are normally associated with rigid body collision detection where the full collision volume can be generated at the pre-processing stage. Pre-processing still occurs in order to generate the individual nodes of the volume hierarchy and these are used as individual rigid bodies in an articulated transformation hierarchy.

An adaptive approach switches between different levels of resolution of the character's volume hierarchy as required, thus facilitating the necessary level of collision detail. As in the approach by Dingliana and O'Sullivan, the volume hierarchy also makes possible the rapid elimination of nodes from the collision detection phase when their parent nodes are found not to be colliding. Due to the simple nature of the nodes in the volume representation, it is possible to incorporate procedures for quickly detecting collisions between the LSS nodes and more general volume representations such as boxes, planes and polygons, which can be used to model the rest of the environment. Thus, it is possible to model not only collisions and self-collisions between characters but also between the characters and a more generic virtual environment.

Once collisions with the humans have been detected, a realistic response is necessary. An optimal solution framework is currently being developed for the inclusion of physically correct responsive objects within the virtual space. This work is primarily concerned with the provision of a feasible real time level of detail hybrid impulse/constraint based solution (see Figure 7). To evaluate our collision handling techniques, psychophysical experiments similar to those described by O'Sullivan and Dingliana²² are being used. We vary the ways in which collision processing can be speeded up and examine the effect of the resultant degradation upon the perception of the viewer.

4. Behaviour

In our system, we allow for several levels of detail with respect to behaviour. Firstly, the concept of *conversational levels of detail* has been used, along with an implementation of *Level of detail AI (LODAI)* for modelling more general behaviour and motivations at multiple levels of detail.

4.1. Conversational LODS

When people come together to socialize and engage in conversations, they become players in an elaborate game governed by many complex rules, both explicit and implicit. These rules are needed to effectively establish and maintain channels of communication among multiple participants. The conversational conduct is carefully coordinated so that, for example, speakers get heard, listener feedback gets seen and topic changes can be negotiated. A large portion of the

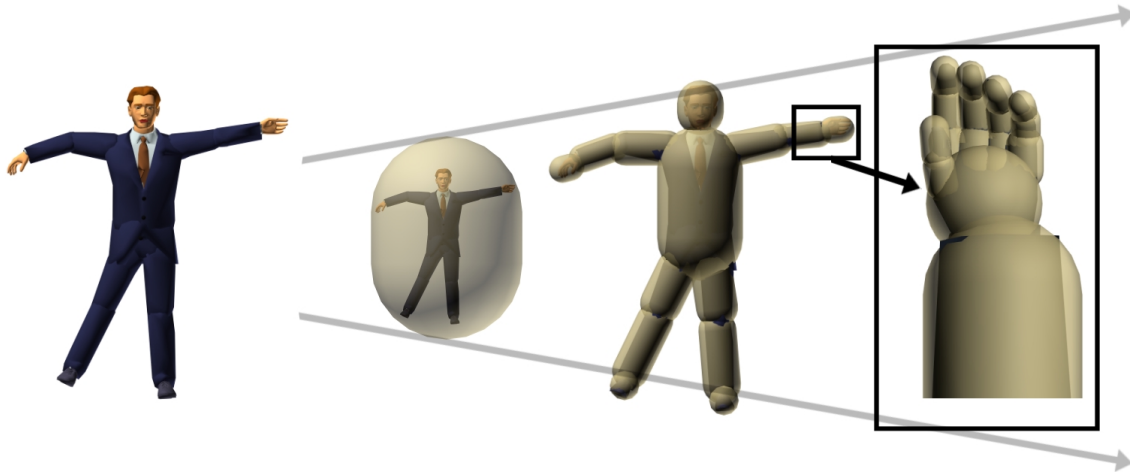


Figure 6: Line Swept Spheres (LSS) are used to efficiently detect collisions with the virtual humans.

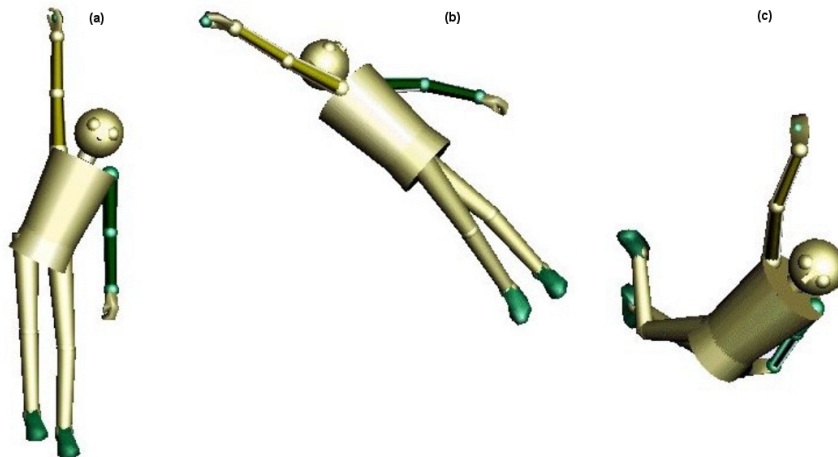


Figure 7: Dynamic constraint-based solution applied to a hierarchical object - (a) figure is resting under gravity; (b) and (c) show external interaction with object - object is being dragged around by the right hand.

coordination is carried out by nonverbal behaviors. For example, those who wish to get the floor often bring their hands up from resting position to indicate their intent, and in turn receive a gaze cue from the current speaker as confirmation that the floor will indeed be turned over to them. While speaking, the speaker only needs to look at a listener and quickly raise their eyebrows to elicit immediate listener feedback such as a head nod ^{11, 12}.

As participants in this game, we rarely have to stop and think about what we are doing, our subconscious social competence keeps us playing along. However, when rules get broken, we quickly spot behaviours that don't fit into the natural flow. Anomalies like that can cause breakdowns in

communication or at least have a jarring effect on participants, even for those that are only observing the event ¹⁴. This has important implications for computer animated social activity, meant to mimic actual human interactions. The animated behaviors cannot break fundamental social rules, or they risk standing out and sending unintended signals of confused activity.

A computational framework has been developed that allows interactive animated characters to have the same properties as humans in face-to-face conversation, including the ability to produce and to respond to both verbal and non-verbal behavior ⁶. Such characters, termed Embodied Conversational Agents (ECAs), have to deal with both multi-

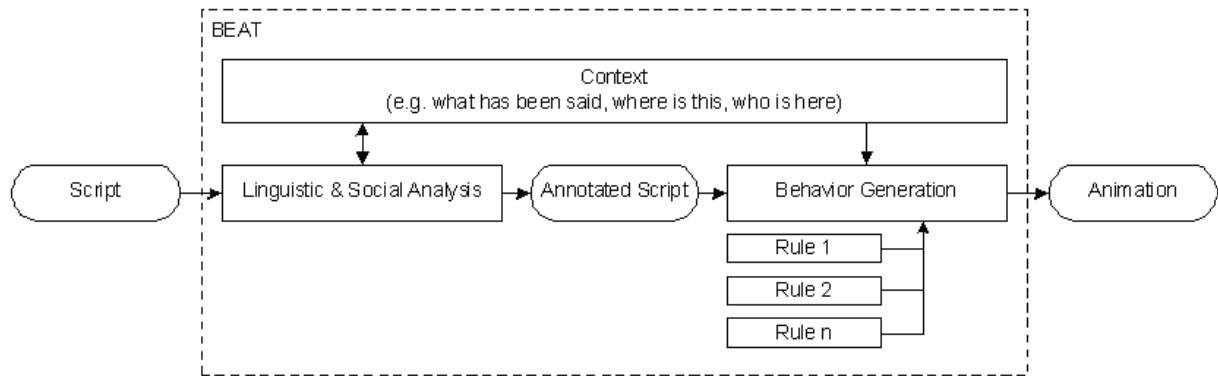


Figure 8: BEAT generates behaviors based on rules that operate on linguistic and social parameters

modal generation and interpretation of conversational content (*e.g.* what is being said) and coordination (*e.g.* taking and giving the floor). Within this framework, a nonverbal behavior generation toolkit called BEAT has been created by Cassell *et al.* ⁵.

BEAT takes as input the text that is to be spoken by an animated human character, and produces as output appropriate nonverbal behaviours closely synchronised with either synthesised or recorded speech (see Figure 9). The input text could be from a prepared movie script, it could be automatically generated by a natural language planning module or it could be extracted from an ongoing conversation among human users of a graphical chat system.

The nonverbal behaviours are assigned on the basis of actual linguistic and contextual analysis of the text, relying on rules derived from extensive research into human conversational behavior (see Figure 8). For example, one rule states that emphasis strokes with a hand or a head nod have a high likelihood of directly preceding or coinciding with a pitch accent, which in turn is commonly associated with new or contrasting spoken material. Such a principled procedural approach guarantees consistency across modalities that is hard to achieve through general stochastic methods and practically impossible to achieve through a purely random behaviour assignment.

The modular nature of BEAT makes it easy to add new rules to generate behaviours as well as rules to filter out behaviours that are conflicting or that meet certain characteristics. For the purpose of animating groups of people having a conversation, BEAT has been extended to include rules for floor management, speaker addressing, feedback elicitation and corresponding listener feedback.

With regard to level of detail (LOD), BEAT lends itself well to controlled generation of behaviours based on visual and functional properties. At the generation level, it is easy to choose which behaviour generation rules are active when processing a given conversation. Filter rules can also be used

to remove generated behaviours based on LOD criteria. At the animation level, since each behaviour generator annotates its generated behaviour with a visual salience parameter, the LOD framework can selectively drop behaviours as the animated character moves further away or out of the focus of attention.

4.2. Level of Detail AI

We are also working on integrating an intelligent agent based role-passing technique into the ALOHA framework. When intelligent agents are used in virtual environments it is often required that they behave believably in a range of different situations. For example, it might be required that within the same simulation an agent is found at work in an office, and then later on enjoying a drink in a bar. See Figure 10 for a typical scenario. The kind of behaviour required of the agent, and the motivations that should drive this behaviour, are quite different in each of these situations. Inspired by (Horswill, 1999) ¹³, the technique of role-passing allows intelligent agents to take on different roles depending on the situation in which they are found.

Role-passing operates by using a schedule to layer appropriate roles on top of a very basic agent at appropriate times within a simulation. This basic agent has a number of attributes which define a character's personality traits, and is capable of simple behaviours such as moving through a virtual world, using objects and interacting with other agents.

When a particular role is layered upon this basic agent, it instructs the agent on how to behave in a certain situation. The first key component of a role is a set of motivations which drive the agent. Activation levels for these motivations are extrapolated from the attributes defining the personality traits of the basic agent. Activation of a motivation results in the agent performing a particular task, such as using a particular object. Secondly, a role contains rules which dictate the agent's interaction with other agents. For example, a role should specify when it is appropriate for an agent to stop and

interact with another character. A role also contains a set of bindings to objects important to the current role. Finally, a role also defines animations required to render the character in that role.

The main advantage of role-passing is the simplicity it lends to populating a virtual world with agents. Placing agents within a novel situation involves simply defining a new role. This eases some of the complications involved in attempting to design very general agents capable of behaving realistically in many situations, and avoids having to write completely separate agents for different roles within a single scene. Another advantage of the role-passing technique is that it moves some way towards creating agents capable of being transferred between different applications. Through role-passing, the same basic agent is able to behave believably in very different situations. This is a major research area in intelligent agent technology ².

Finally, role passing allows the use of Level of Detail Artificial Intelligence (LODAI). By assuming and discarding roles as required, motivations unrelated to the current situation encountered by an agent are never considered. The result of this is that motivation levels surplus to the current situation need not be stored and decisions not related to the current situation are never even considered. Additionally, LODAI techniques can be further utilised when defining roles by specifying an agent's behaviour at varying levels of detail. For example, if the agent is not in the current view frame, behaviours required only to make the agent appear believable are omitted and only those behaviours crucial to the simulation need to be included. This avoids updating complex behaviours for less important characters, while still maintaining some basic functionality as they may become more salient later. See MacNamee *et al.* ¹⁸ for further details.

5. Conclusions and Future Work

This paper has presented ongoing work on a framework for achieving level of detail human animation for achieving crowd and group simulation in real-time. This allows refinement of the animation both at a macro level, whereby full components of the system may be activated or deactivated based on importance heuristics, and within the individual components themselves. At the geometrical level, subdivision techniques can be used to achieve smooth rendering LOD changes, while other objects can be pre-emptively simplified. At the motion level, the movements themselves can be simulated at adaptive levels of detail. For behaviour, LODAI can be employed to reduce the computational costs of updating the behaviour of characters that are less important. Purely random playback can be used for characters that are not important to the scene, while motions that are synchronised with speech, using the BEAT framework, can be played for more salient beings. Furthermore, the knowledge embedded in the system can be used to allow the gesture generation engine to make informed decisions about which

features are most important to retain based on the salience of an agent or group of agents, thus allowing graceful degradation of the gesture repertoire.

Although significant gains can be achieved by using level of detail techniques for all tasks involved in the simulation of a complex graphical environment, in-built redundancies could be introduced if the system fails to share the prioritisation information generated by one process with the others. We want to make the LOD resolver more generic, so that truly polymorphic LOD control can be achieved. This will involve developing a seamless interface to the knowledge base that can be used to schedule the processing of each of the constituent components of the system.

Acknowledgements

This work has been supported by Enterprise Ireland and the Higher Education Authority of Ireland.

References

1. A. Aubel, R. Boulic and D. Thalmann. "Real-time Display of Virtual Humans: Level of Details and Imposers", *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on 3D Video Technology* **10**(2), pp. 207–217 (2000). [2](#)
2. R. Aylett, K. Dautenhahn, J.E. Doran, M. Luck, S. Moss and M. Tennenholtz. "Can Models of Agents be Transferred between Different Areas?", *Knowledge Engineering Review*, **15**(2), pp 199–203 (2000). [7](#)
3. D. Baraff and A. Witkin. "Large Steps in Cloth Simulation", *SIGGRAPH 98 Conference Proceedings*, pp. 43–54, 1998. [4](#)
4. D.C. Brogan and J.K. Hodgins. "Group Behaviors for Systems with Significant Dynamics", *Autonomous Robots* **4**, pp 137–153 (1997). [1](#)
5. J. Cassell, H. Vilhjálmsón and T. Bickmore. "BEAT: Behavior Expression Animation Toolkit", *Proceedings SIGGRAPH 2001*, pp 477–486 (2001). [6](#)
6. J. Cassell, H. Vilhjálmsón, K. Chang, T. Bickmore, L. Campbell and H. Yan. "Requirements for an Architecture for Embodied Conversational Characters", *Computer Animation and Simulation '99 (Eurographics Series)*, Thalmann, D. and Thalmann, N. (eds.), pp. 109–120, (1999). [5](#)
7. T. DeRose, M. Kass and T. Truong. "Subdivision Surfaces in Character Animation", *Proceedings SIGGRAPH 1998*, pp 85–94 (1998). [2](#)
8. J. Dingliana and C. O'Sullivan. "Graceful Degradation of Collision Handling in Physically Based Animation", *Computer Graphics Forum. (Proc. Eurographics 2000)* **19**(3), pp 239–247 (2000). [4](#)

9. N. Dyn, J.A. Gregory and D. Levin. "A Butterfly Subdivision Scheme for Surface Interpolation with Tension Control", *ACM Transactions on Graphics* **9** (2) pp 160–190 (1990). [2](#)
10. T. Giang, R. Mooney, C. Peters and C. O'Sullivan. "ALOHA: Adaptive Level Of Detail for Human Animation: Towards a new framework", *Eurographics 2000 short paper proceedings*, pp 71–77 (2000). [2](#)
11. E. Goffmann *Behavior in public places; notes on the social organization of gatherings*, New York, Free Press of Glencoe, (1963). [5](#)
12. C. Goodwin *Conversational Organization: Interaction between speakers and hearers*, New York, Academic Press, (1981). [5](#)
13. I.D. Horswill and R. Zubek, "Robot Architectures for Believable Game Agents", *Proceedings of the 1999 AAAI Spring Symposium on Artificial Intelligence and Computer Games, AAAI Technical Report SS-99-02*, (1999). [6](#)
14. A. Kendon. *Conducting Interaction: Patterns of behavior in focused encounters.*, New York, Cambridge University Press, (1990). [5](#)
15. E. Larsen, S. Gottschalk, M.C. Lin, and D. Manocha. "Fast proximity queries with swept sphere volumes", *Technical Report TR99-018, Department of Computer Science, UNC Chapel Hill* 1999. [4](#)
16. W. Leeson. "Subdivision Surfaces for Character Animation", *Games Programming Gems III*, pp 372–383 (2002). [2](#)
17. C. Loop. "Smooth subdivision surfaces based on triangles", *Master's thesis, University of Utah, Department of Mathematics* (1987). [2](#)
18. B. MacNamee, S. Dobbyn, P. Cunningham and C. O'Sullivan. "Men Behaving Appropriately: Integrating the Role Passing Technique into the ALOHA System", *Proceedings of the AISB'02 symposium: Animating Expressive Characters for Social Interactions (short paper)*, 59-62, (2002). [7](#)
19. S.R. Musse, F. Garat and D. Thalmann. "Guiding and Interacting with Virtual Crowds in Real-time", *Proceeding of the Eurographics Workshop on Computer Animation and Simulation (CAS '99)* pp 23–34 (1999). [1](#)
20. S.R. Musse and D. Thalmann. "A Behavioral Model for Real Time Simulation of Virtual Human Crowds", *IEEE Transactions on Visualization and Computer Graphics* **7**(2) pp 152–164 (2001). [1](#)
21. Musse, S.R. and Thalmann, D. "A Model of Human Crowd Behavior: Group Interrelationship and Collision Detection Analysis", *Proc. Workshop of Computer Animation and Simulation of Eurographics'97*, Sept, 1997. Budapest, Hungary.
22. C. O'Sullivan and J. Dingliana. "Collisions and Perception", *ACM Transactions on Graphics* **20** (3) pp 151–168 (2001). [4](#)
23. C. Peters and C. O'Sullivan. "Vision-Based Reaching for Autonomous Virtual Humans", *Proceedings of the AISB'02 symposium: Animating Expressive Characters for Social Interactions (short paper)*, 69–72, (2002). [4](#)
24. C. Reynolds. "Flocks, Herds, and Schools: A Distributed Behavioral Model", *Proceedings SIGGRAPH 1987* pp 25–34 (1987). [1](#)
25. F. Tecchia, C. Loscos and Y. Chrysanthou. "Real Time Rendering of Populated Urban Environments", *ACM Siggraph'01 technical sketch* (2001). [2](#)
26. F. Tecchia and Y. Chrysanthou, "Real-Time Rendering of Densely Populated Urban Environments", *Rendering Techniques'00 (Proceedings of the 10th Eurographics Workshop on Rendering)*, pp 45–56, (2000). [2](#)
27. D. Zorin, P. Schröder and W. Sweldens. "Interpolating Subdivision for Meshes with Arbitrary Topology", *Proceedings SIGGRAPH 1996*, pp 189–192, (1996). [2](#)

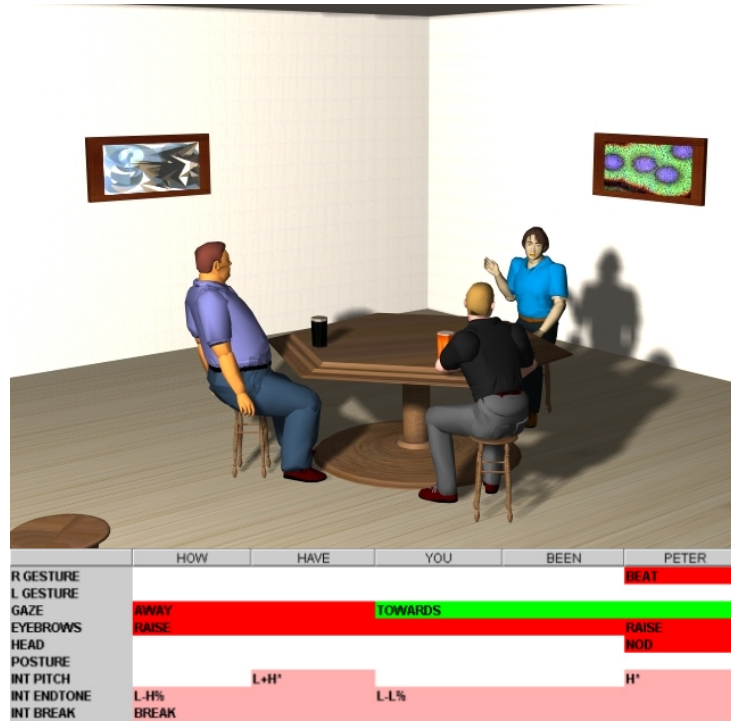


Figure 9: The BEAT gesture toolkit is used to generate speech synchronized output for the speaker.



Figure 10: A group of agents drinking in a bar - suitable roles must be chosen for each individual.