Attack on the Clones: Managing Player Perceptions of Visual Variety and Believability in Video Game Crowds

Sean Oxspring, Ben Kirman and Oliver Szymanezyk

Lincoln Games
Research Group
University of Lincoln, UK
{soxspring,bkirman,oszymanezyk}@lincoln.ac.uk

Abstract
Crowds of non-player characters are increasingly common in contemporary video games. It is often the case that individual models are re-used, lowering visual variety in the crowd and potentially affecting realism and believability. This paper explores a number of approaches to increase visual diversity in large game crowds, and discusses a procedural solution for generating diverse non-player character models. This is evaluated using mixed methods, including a "clone spotting" activity and measurement of impact on computational overheads, in order to present a multi-faceted and adjustable solution to increase believability and variety in video game crowds.

Introduction
Procedural crowd generation techniques are increasingly common in the video games industry to simulate large groups of agents. Hordes of zombies have been rendered in games such as the *Left 4 Dead* series (Valve, 2008 and 2009) and *Dead Rising* (Capcom, 2006). Games such as *Hitman Absolution* (IO Interactive, 2012) and *Assassins Creed 2* (Ubisoft, 2009) also rely on the simulation of "realistic" and "believable" multi-agent crowds for the player to interact with.

When simulating crowds in a virtual environment such as a game, it is often hard to produce a great deal of variety in characters. Usually due to the complexity of manual character modelling, it is very time-consuming for a 3D modeller to manually create a

great deal of heterogeneous agents [0]. It is common to use a fixed number of template characters to instance into an environment; this, however, may lead to the problem of visually similar 'appearance clones' in the environment [0]. To combat this, procedural generation of characters is becoming a more popular solution [0]. This paper explores techniques in which crowd generation algorithms can be adjusted to better support greater diversity, in order to support more believable play environments. Firstly, a case study of an approach used in a commercial game series is explored, and then we introduce a new, multi-faceted approach to crowd generation, and evaluate it in terms of computational overhead and effect on perception of diversity.

Commercial Approaches to Crowd Variety

Procedural Content Generation (PCG) involves the generation of game content in real-time. Randomness is often a key part of procedural generation, allowing a broad range of possibilities to be generated from a small number of parameters. Many games use PCG to generate levels and unique gameplay. *Borderlands* (2K Games, 2009) uses PCG to generate thousands of permutations of guns. *Minecraft* (Mojang, 2011) generates enormous worlds using a procedural algorithm. PCG is also widely applied to the generation of human agents in real-time for use in a crowd of non-player characters (NPCs) in games. Following a simple generation algorithm, hundreds of thousands of unique-looking character models can be created in real time.

Case Study: *Left 4 Dead* and *Left 4 Dead 2*

Left 4 Dead (Valve, 2008) is a cooperative first-person shooter in which players must fight their way through hordes of infected humans during an apocalyptic pandemic. The game uses an artificial intelligence system known as 'The Director' in order to promote replay value and achieve a sense of "Dramatic Game Pacing" [0].
As a zombie game, "infected" characters were designed to be "visible but not memorable" since they appear en masse to challenge the players. Developers reported that one of the early issues was that through lack of diversity, the reuse of models meant that a small set of infected characters became more memorable and familiar. Based on this, additional variation systems were developed [0].

Limited Sharing of Head Textures: Parts of textures would be merged together in order to create a finalised head texture. This would include mouths, eyes and other facial features (see Fig 2)*.* The possible combinations of different features greatly increased potential diversity in absolute terms.

Fig. 1. Left 4 Dead Male Zombie Head Variants

Colour Tinting on Clothes: The clothes on common infected were tinted using a random multiplication of RGB values. This multiplication allowed for the clothing texture to maintain its detailing, as contrast was never lost. Dark parts of the clothes would remain dark no matter what was generated.

Grimes [0] states that fine details and low-contrast areas on textures did not work well to create visual variety amongst members of the horde, as these were difficult to change using colour multiplication techniques and lead to poor-quality textures on models. She declares that geometric variation on the models warped facial features and therefore helped to hide re-used textures. The image below shows two different models with the same facial texture and how they both look visually different from one another:

Fig. 2. Left 4 Dead Facial Geometric Variation and Texturing

Decals such as blood flecks and dirt were also very effective in decreasing the amount of times that playtesters noticed clones of recognisable - the decals obscured recognisable features and allowed one texture to be re-used multiple times without players noticing during play.

Left 4 Dead 2 (Valve, 2009) is the sequel to *Left 4 Dead*. As with the original, it was a first person shooter with an emphasis on four-player teamwork in order to complete missions set in a zombie apocalypse scenario. Focus was put on expanding the game experience and making it more complex than the predecessor. This included the development of an improved system for generating varied common infected [0]. A new shader was developed in order to create even more variance in the common infected models. According to Vlachos [0], the simplest infected has over 24,000 variations. This large amount of visual variation were achieved by a combination of the following methods:

Gradient Mapping in the Luminosity channel was used to tint the textures on the common infected models. It was noted that one single gradient map could not be used for the whole model as having more than one colour in the gradient map caused problem with inappropriate colouration. Grimes [0] describes that the problem was overcome by exclusively masking certain areas of the texture, allowing the tints to be applied to exclusive areas of the final texture.

Blood and Grime Decals were added to the models in a similar technique to that of *Left 4 Dead.* Blood masks were created and were then added on top of the characters tinted texture. Grime was then added on top in the same fashion; since grime did not need to be as defined as blood, this was added using transparencies and a high detail texture [0].

Player Perception of Variety in Commercial Games
In order to better understand the perception of diversity in crowds in games from a player's point of view, a pilot study was conducted. Three groups of randomly chosen people (N=14) were shown two videos of gameplay from commercial games that include crowds. The groups were told to pay particular attention to the crowds, and less about the game itself. They were then asked about what they noticed about the crowds. Gameplay was recorded from *Hitman Absolution* (IO Interactive, 2012) and *Assassins Creed 2* (Ubisoft, 2009). Levels used were specifically chosen to have large crowds of non-player characters.

Fig. 3. Non-Player Crowds in *Hitman: Absolution* (left) and *Assassins Creed 2* (right)
In unstructured discussions, participants shared their thoughts on the crowds in these specific titles. For *Hitman: Absolution*, participants were generally positive about crowd variety, however they did often see repetition of clothing on the models "*I think there were a couple of models that were there a couple of times*". The crowds of people were also described as "*dull and generic*" which was identified as drawing the eye away from looking at the crowd too closely.

For *Assassins Creed 2*, participants were much more negative, repeatedly pointing out the use of the same models for clothing on the characters. They suggested that the game merely manipulated the colours of the clothes and left the models the same. One member of a group pointed out that "*Everyone was the same height*". All three groups were quick to point out the use of similar if not identical clothing models on groups of people such as a group of monks in the video "*[the] attack of the clone monks*". Two of the groups picked out jarring movements and collisions, where characters in the game would crash into each other randomly and not avoid each other correctly.

In both games, participants pointed out noticing motion clones (characters doing the same/similar motion) during the videos. They also described motion of characters in the *Hitman: Absolution* crowd as being 'unrealistic' as they would often pace around randomly with no clear direction.

In all focus group sessions, there was a strong suggestion that appearance clones would impact negatively on the player's sense of believability (and therefore immersion) whilst playing a game.

This activity identified the two major factors affecting believability in these crowds were appearance and movement. This supports by findings of previous work on crowd diversity by McDonnell *et al.* [0] that also suggests that the visual appearance of the characters is more important in creating perceived diversity than their motion.

A Procedural Approach to Diversity

As underlined in the previous sections, visual diversity remains a problem in video game crowds. The examples have shown how existing games have approached the task of creating variation. Based upon the suggestions from the focus group, the approaches from the games industry and proposed variety features in the works of Thalmann and Musse [0],.a list of several visual variety-changing factors were collated for inclusion in a proposed multi-dimensional solution. All of these changes are focussed on addressing the variety of appearance of non-player characters in the crowd. Each approach focuses on altering the virtual representation of the game character by altering either the shape of the character model or the post-processing effects used on each character.

Colouring

One of the simplest approaches to increase visual diversity is through the use of random colouring to different aspects of the character model (i.e. clothing, skin tone, hair). RGB values can be randomly generated and applied to various parts of the character, with the use of a Colour Look Up Table (CLUT) in order to limit potential colours to realistic subsets and distributions. However, Thalmann and Musse [0]. point out that even with techniques of colour change, it does not overcome the feeling of

watching the same model - thus physical model changes must be considered as more important than colour changes.

Body Somatotypes

Somatotype theory is a well-understood framework for the description and categorisation of human body size and shape based on three axes: the ectomorph, mesomorph and endomorph [0]. Each element is defined by a number between 1 and 7, which contribute to a finalised somatotype combination.

Although Somatotypology was originally used in now discredited work on connecting body shape and disease-susceptibility, this system is still useful as a model for generating humanlike figures. Basing the procedural generation of character models on this kind of model ensures that the final models follow shapes that are found in the general population. Implementation of these requires model vertices that allow for deformation of models based on the somatotype elements assigned. An example of this would be deforming a convex stomach area for an endomorph-like character or a concave stomach for an extreme ectomorph.

In addition, character heights can be implemented alongside somatotypes using scaling methods, although careful attention must be made to ensure that the head of the characters is not warped to anomalous dimensions. Additional weight could also be simulated with this method by scaling the characters in the x and z axis.

Accessories

The addition of possible items such as hats and glasses to characters can also potentially increase variety. Depending on gender, beards, moustaches and other (generally) gender-associated accessories could be added to character models. These accessories would be modelled separately and then instantiated and rendered using the same world transformation as the character that they are assigned to.

Model Design

The design of the basic/archetypical individual character model itself is a very difficult issue tied to realism and believability. We argue that the believability of crowds is independent of the aesthetic context within which those crowds exist. Specifically, believability is not necessarily the same as realism. Where realism is concerned with the simulation of the real world as accurately as possible, believability is about maintaining a situation that is consistent with the player's expectations within the "magic circle" of the game. For example, it is not realistic that characters in *Super Mario Bros* (1985, Nintendo) exist only within a 2-dimensional space where mushrooms have magical abilities. However, in this context, it would be believable for the consumption

of a mushroom to cause a character to greatly grow in size. In contrast, it would be unbelievable if Mario were to move into a third dimension, since (in this title) this behavior is inconsistent with the encompassing game fiction. In terms of visual aesthetics, a consistent visual style is similarly important to maintaining a sense of believability. The massive success of *Minecraft* (2011, Mojang), with its distinctive blocky style, has demonstrated that consistent visual aesthetics can create engaging and consistent (believable) experience [0] in spite of its lack of visual realism.

Based on this argument, the design of model used to demonstrate and evaluate the variety of methods for increasing diversity was specifically chosen to be fairly blocky. In addition, the character models feature no animation, since this research is focused purely on visual diversity rather than movement diversity (as discussed in [0]). Within the fictional world of the test game used to evaluate this methodology, although lack of animation is not realistic, it is believable and consistent with the visual aesthetic.

The model itself is stored within a vertex array. It consists of three cubes with their bottom faces removed (as they do not need to be rendered)*. The model also makes use of back-face culling, a process in which the back of polygons are not rendered.

The models texture coordinates are stored within another vertex array and are then used to map the texture to its appropriate location in 3D space. Surface normals are also stored in a separate array. This basic model has been extended with extra vertices on the torso in order for deformation of the model to occur. The extra vertices are designed to be moved from their neutral position based on randomly selected somatotype values. When put together these three values and the neutral vertices will produce a complete humanoid body shape. Figure 6 shows the three extreme body types as they would look in the newly designed model.

The *Ectomorphy* value influences the stomach and sides of the character's torso, giving the model a leaner appearance. The *Endomorphy* value affects the vertices in the opposite way, pushing the model's vertices outwards and giving it a more rotund appearance. The *Mesomorphy* value also dampens the effect of the *Endomorph* value on the character model's stomach. It primarily pushes the model's shoulders outwards to create a broader looking model. Examples are shown in figure 6.

The model is therefore capable of being generated based on deformations to model following this approach, and the inclusion of random colour mapping for clothes, hair and skin-colour, height/weight scaling and random accessories as required.
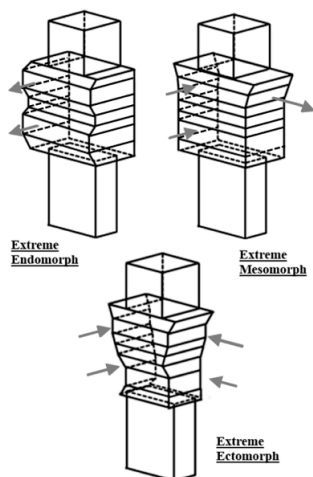
Fig. 4. Extreme Somatotypes and deformations on the character model vertices

Evaluation

The multi-faceted nature of the proposed approach to generating diversity creates the opportunity to evaluate the different approaches to introducing variety based on their effect on believability. In addition, it is also important to look at the processing cost of the various approaches, as this will help to evaluate whether or not the changes to a complexity of procedural character generation system are cost-effective in terms of increasing perceived diversity whilst not impacting too heavily on overhead.

Perception of Variety Evaluation Methodology

Given the argument that lack of variety is associated with negative believability (and therefore engagement/immersion in a game world), we built the evaluation to test player perception of variety in crowds.

Crucially, it is the player's perception of variety that is important, instead of some automated measure of diversity based on combinations of generated models. Although a real crowd has vast diversity based on the millions of differences between individuals, the combination of factors that make a virtual crowd believable may be a much smaller subset. For example, many of the more subtle differences between individuals may have different value in the perception of diversity.

Indeed, it is also exceedingly important that false positives (i.e. player incorrectly identifies a false lack of diversity) are taken into account since it is user perception that is most important (more so than realism, for example). Increasingly subtle changes to the procedural generation of crowds may have no meaningful effect on increasing perceived diversity (i.e. believability).

McDonnell et al. [0] conducted an experiment into the perceptual impact of appearance and motion "clones" in simulated crowds. In their experiment participants were asked to watch a crowd simulation running and then click on clones of characters as and when they noticed them - the time between the start of the simulation and the click was then recorded.

Using elements from the technique described from above, we performed an evaluation to investigate the effectiveness of the variety of factors in the proposed crowd generation algorithms. This does not look at the speed at which participants find clones - instead it looks at the total clones that the player notices within a certain time period as a measure of player perception of variety.

Procedure

Fifty naive individuals between the ages of eighteen and thirty (26 Male, 24 Female) were asked to take part in the evaluation. They had mixed educational backgrounds and not all had an interest in computer gaming. Each participant was required to provide informed consent and retained anonymity in the study.

Participants were then asked to sit down and watch one of five pre-recorded video sets. These sets of videos had the same six fifteen-second videos in them, but in a randomised order to avoid unintended variables.

Each video featured an animation of a crowd of non-player characters in a virtual test environment (described in [0]), where the generation of character models in each condition used a different combination of approaches. The videos were screen-captures of the crowd simulation in action, but with differences to the way in which the character models were generated.

All videos contain the crowd moving around the environment at an average walking pace as this better recreates a situation in which the crowd would actually be viewed (instead of having all the characters just standing still). The camera was also positioned at eye-level with the other characters in order to simulate the height at which the crowd would generally be viewed. A fixed seed was used to ensure crowds behaved identically between participants.

Participants were individually shown the six videos in random order. They were asked to take a tally whenever they saw two agents in the video that they considered to look the *same.* In this way, perception of variety is recorded based on the number of clones identified, across conditions to allow comparison of the facets of the crowd generation technique.

Results

Data collected represent the inverse of the user's perception of diversity in the crowd simulation. The more clones that were spotted suggest less perceived diversity in the crowd regardless of how many "real" clones there were. False positives still lower the perceived diversity in the same way true appearance clones. The population mean calculated from the results suggests that as the models became more varied, it became harder for participants to pick out characters that looked the same within the scene. This supports the case that the improvements made to the basic character generation system (i.e. fixed set of model skins) have been effective; reducing the average of 6.48 clones spotted per session to 3.1 clones.

Table 1. Comparison of clone perceptions across conditions

	n Methodology	nes Spotted (s.d.)
	entical models, skin variants	)
	on 1, with height/weight scaling	)
	matotype generation	)
	on 3, with colour tinting	)
	on 4, with model accessories	
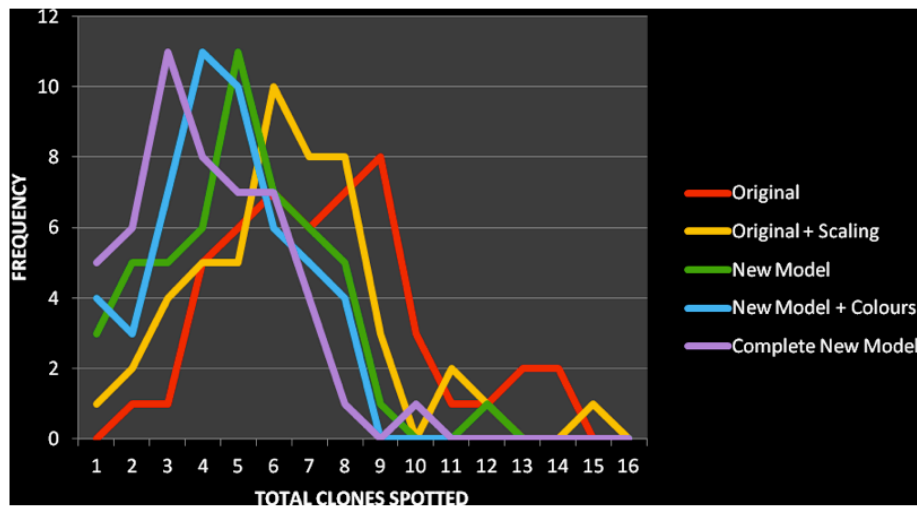
Fig. 5. Distribution of clones identified between conditions

Figure 7 shows the frequency of total clones spotted in each condition. As the model's variation increases, the graph peaks tend towards to lower totals. This shift towards the left implies that it was harder for participants to notice appearance clones as variation increased in the scene.

It is also interesting to note that there is a flatter distribution in the 'original' system's results compared with that of the more complex systems. This increased spread may be due to the inability of some participants to be totally sure that two figures looked

entirely the same. As the variation increased it may have been easier for test-subjects to define similar-looking characters due to their more obvious colour changes and similar accessories.

It also appears that basic scaling on the original model (the yellow line on Figure 7) had a dramatic difference to that of its predecessor. This suggests that merely adding the random scaling to each character could provide the greatest benefit on the crowd whilst also being the lowest cost on processing power (compared with the other changes to the model).

It was also noticed that participants spotted many false-positives in the final model version. This may be due to the addition of the accessory hats, which may have made the models look more visually similar to one another than without them (i.e. two men with top hats may have been considered as appearance clones even when they had totally different heads or bodies.) This is an interesting result since it shows that the use of limited accessories such as hats or bags in procedurally generated crowds should be done with great caution since they act as a signifier of lack of diversity. Unpaired t-Tests were conducted to compare the data collected from the original model test with that of the data collected from the other models. This was done to see whether there was a chance that the models did not actually affect the perceived diversity in the scene. The comparison between the basic (random skin) model and others showed a significant difference in the number of clones identified (therefore perceived diversity) ($p<0.001$) in each condition, except the basic scaling version where $p <0.05$.

In order to see how the extra complexity of additions to the new model (colours and accessories) affected the perceived diversity, t-tests were conducted on the new model and its more complex variations (colour palette alterations and accessories). The results show that while variation of colours was not significant in increasing perceived diversity ($p=0.26$), the addition of random accessories did significantly increase diversity ($p<0.05$).
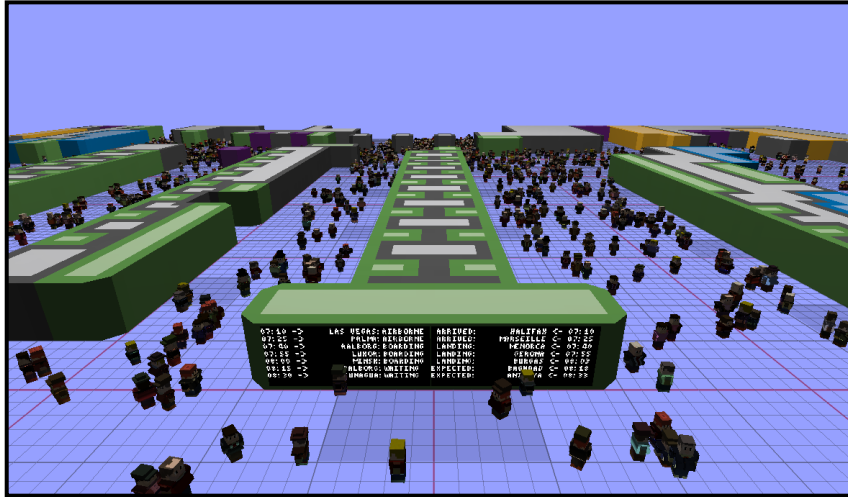
Fig. 6. High Populations of Agents in crowds (1400+ individuals)

Model Complexity and Computational Overhead

Having a high frame rate per second (FPS) is an important factor in the fidelity of graphics software and games [0]. Therefore, it was important to analyse the effect that the new model had on the program's frame rate as the number of agents increased. An experiment was run in which the FPS of the software was recorded as the total number of agents in the scene increased. An agent was instantiated once per frame until reaching a maximum of 1,500 agents at once. Data was recorded once every three frames to reduce the load on the system (see figure 9).

Four tests were carried out based on four conditions: The basic model shape, the "best case" of the new model (lowest possible polygon count), "average case" (normal probability of additional accessories/polygons), and worst cast (all agents have maximum polygons allowed by accessories).

In order to ensure a fair test, the four simulations were run on the same machine with a minimised number of other programs active at run-time. The simulations were carried out using a locked random seed, ensuring that the movement and character generation would be the same each time (the same as with the previous experiment). Comparing the *average case* of the new model with the *original* model's frame rate shows an average difference of 1.568132 frames per second which at high frame rates is negligible, however the frame rate for the new model drops below twenty-five frames per second at around 376 total agents whereas the frame rate for the original drops at 413, which is a noticeable difference of 37 extra agents.

There is also a dramatic difference between frame rate for the original model and the worst-case scenario (in which every member of the crowd simulation is wearing a hat) however the probability of this case occurring during the simulation is incredibly slim (P< 0.001 based approximation using Normal Distribution).

Although the new model does have a slightly lower frame rate than that of the original model, the increased processing overhead is worth the trade-off considering its noticeable effect on the viewer's perception of the crowd (52.16% improvement on the original in terms of participants perception of appearance clones).

Conclusion

This paper has investigated the player perception of diversity in video game crowds. In particular, where crowds are procedurally generated, there is a potential gameplay issue where players experience loss of immersion when exposed to game crowds featuring frequently repeated character models. This was confirmed through a pilot qualitative focus group activity where participants watched game videos and commented on lack of diversity in crowds.

Taking inspiration from Sheldon's body somatotype theory a new character generation algorithm was designed with a character model mesh that would deform to represent different body shapes [0], and also vary based on colour, height, weight and accessories (e.g. hats). In experimental conditions, the different facets of the model generation process were tested to understand user *perception* of diversity in virtual crowds. This relative measure is much more important than the absolute measure of diversity since it is based on the personal experience of the player within the game.

With all character generation aspects operating, the new model significantly improved upon the basic version (P < 0.0001) in terms of reducing the user's perception of crowd diversity. However, we identified that when the new model was colour-tinted there was no noticeable statistical difference in user perception of diversity. This suggests that colour tinting may not be as effective as previously thought, or that the differences in colours were not vibrant enough for the user to notice.

The new model generation algorithm appears to have reduced the user's perception of appearance clones within the scene. In further tests of computational overhead required by this multi-faceted approach, we highlight the cost/benefit comparison of the approach based on best- and worst-case scenarios, showing how the greatest perceived crowd diversity can be achieved for lowest computational cost.

This work focussed explicitly on appearance clones on moving but not animated models. Although this fits within the visual aesthetic of the game environment chosen (which has a lo-fi style), in other more realistic environments the movement and visual detail of models will no doubt have an effect on user perceptions of clones. The specific

factors that are important for different aesthetic styles will vary and as such some aspects of the study may have less or more importance depending on the presentation of a specific game.

References

Booth, M (2009). *The AI Systems of Left 4 Dead*. (pdf). Available from: <http://www.valvesoftware.com/publications/2009/ai_systems_of_l4d_mike_booth.pdf>.

Claypool, M. Claypool, K. Damaa, F (2006). *The effects of frame rate and resolution on users playing First Person Shooter games*. Electronic Imaging 2006.

Grimes, B (2010). *Shading a Bigger, Better Sequel.* (pdf). Available from: <http://www.valvesoftware.com/publications/2010/GDC10_ShaderTechniquesL4D2.pdf>. [Accessed 2 January 2013]

Kirman, B. (2010) Emergence and Playfulness in Social Games, in Proceedings 14th Academic Mindtrek Conference, Tampere, ACM.

Lewis, M (2000). *Initial Models created by Lewis's Maya Plugin* [online]. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.3410>. [3/1/2013].

McDonnell R, Larkin M, Dobbyn S, Collins, O'Sullivan C (2008). *Clone attack! Perception of crowd variety*. In: SIGGRAPH 2008, August 2008, Los Angeles. Article No. 26.

Sheldon W., (1970). Atlas *of Men: A Guide for Somatotyping the Adult Image of All Ages*. New Edition. Macmillan Pub Co.

Szymanezyk O, Dickinson P and Duckett T (2011) *Towards agent-based crowd simulation in airports using games technology. In: Agent and multi-agent systems: technologies and applications*. Lecture Notes in Computer Science, 6682 (6682). Springer-Verlag, Berling Heidelberg, pp. 524-533.

Thalmann, D and Musse, S. R (2000). *Somatotype Deformation on a Simple Human*

Thalmann, D and Musse, S. R (2007). *Crowd Simulation*. 1. ed. London: Springer.

Vlachos, A (2010). *Rendering Wounds in Left 4 Dead 2.* (pdf). Available from: <http://www.valvesoftware.com/publications/2010/gdc2010_vlachos_l4d2wounds.pdf>. [Accessed 2 January 2013]