

Perceptual Evaluation of LOD Clothing for Virtual Humans

R. McDonnell, S. Dobbyn, S. Collins and C. O'Sullivan

Interaction, Simulation and Graphics Lab, Trinity College Dublin, Ireland

Abstract

Recent developments in crowd simulation have allowed thousands of characters to be rendered in real-time. Usually this is achieved through the use of Level of Detail (LOD) models for the individuals in the crowd. Perceptual studies have shown that image-based representations, i.e., impostors, can be used as imperceptible background substitutes for high-polygon models for skinned human characters, resulting in optimal rendering times and high visual fidelity. However, previous methods only showed humans dressed in clothes that were deformed using standard skinning methods. Highly deformable objects like cloth are not effectively depicted using these methods. Therefore, in this paper, we present the first perceptual evaluation of different LOD representations of humans wearing deformable (i.e., physically simulated) clothing. We show conclusively that impostors are startlingly effective at depicting the deformation properties of clothing and present useful guidelines for the development of crowd systems with thousands of realistically clothed humans.

1. Introduction

Displaying large crowds of high quality virtual characters with deformable clothing is not possible in real-time at present because of the cost of rendering the thousands of polygons necessary to depict the subtle motion of the cloth. Current real-time crowd systems are capable of displaying thousands of skinned characters by using lower quality representations instead of high quality to achieve their goal. Hybrid systems that switch between high and lower quality models depending on the distance from the camera, are also a solution to this problem.

In this paper, we assess the effectiveness of different LOD cloth representations, in order to find which (if any) could be displayed seamlessly alongside the higher quality cloth in large crowds. We feel that crowds of characters with deformable clothing could be possible, if one of these techniques proves effective. Therefore, we present the first psychophysical analysis of simulated cloth on the most popular human representations used in crowds; geometry and image based representations (i.e., impostors), in order to determine if these representations are suitable for displaying the subtle folds and material properties of cloth that are very important for realism.

With this in mind, we address the questions: Can impostors and low resolution geometry display a range of different cloth materials? How well can they reproduce individ-

ual material types? If compared directly, which representation would resemble a higher quality representation more closely? What is the best update frequency for impostor images? What happens when viewing a crowd of these representations?

2. Background

Recent methods for rendering real-time crowds involve replacing complex virtual human models with simpler representations, such as lower resolution models containing fewer polygons [dHCUCT04] and impostors consisting of camera-facing quadrilaterals texture-mapped with images of a human model [TLC02] [DHOO05].

Previous studies on these simpler representations has focused on the perception of the appearance and motion of characters clothed with skinned meshes. Hamill et al. [HMDO05] perceptually evaluated geometry and impostors for buildings and humans and gave metrics on the distances at which impostor representations are indistinguishable from high resolution geometry, for a hybrid geometry-impostor switching system. McDonnell et al. [MDO05] performed a set of psychophysical experiments to test the differences in motion perception when viewing different LOD human models. They found that participants were more sensitive to changes in motion when viewing an impostor than when viewing a low resolution polygon model. They concluded that subtle motion information was being lost due to

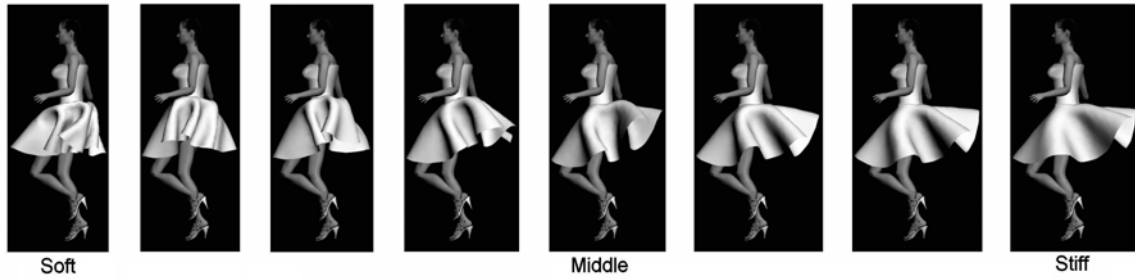


Figure 1: Eight different materials used in order of increasing stiffness

the reduction in polygons in the low resolution model and suggested that the impostors retained deformations in the models better because they were generated from a higher resolution polygon model. If this was indeed the case, we hypothesize that impostors should therefore be particularly suited to representing highly deformable cloth simulations.

There is no existing evidence that impostors are a valid substitute for high resolution geometry cloth simulation - previous perceptual experiments only showed that they were effective for normal walking skinned characters, but not characters with physically-simulated, highly deformable clothing. Other crowd simulation methods use low resolution geometry effectively for large crowds (albeit with skinned, not physically simulated clothing). There is also no evidence that low resolution geometry would be suitable for the depiction of deformable cloth. In our experiments, we test observer perception of different animations of cloth with varying properties. We wished to examine how similar the perception of cloth deformation with impostors or low resolution geometry is to that of the high resolution polygonal cloth simulation.

There has been no previous work to date on analyzing the perception of cloth motion. However, a perceptual metric was used by Bhat et al. [BTH*03], based on neurobiological studies on the human perceptual system's sensitivity to edges or lines. They used this metric to compare video of real cloth with simulation and thereby devised a new algorithm to determine parameters for different fabrics.

3. Psychophysical Experiments

Our aim was to create a hybrid crowd system similar to that described in [DHOO05] but with clothed characters. However, we were not certain at the outset which representation (i.e., low detail geometry or impostor) would be most suited to displaying the deformations of simulated cloth, although we hypothesized that low geometry would not be sufficient to reproduce the required deformations.

We performed a set of six psychophysical experiments in order to provide guidelines to assist our system:

1. Determining Perceptually Linear Scale

Aim: To find a suitable graduated scale of cloth properties so that we could methodically select fixed cloth motion steps.

2. Stiffness Sorting Experiment

Aim: Assess the ability of low detail geometry, high detail

geometry and impostors to visualize cloth effectively.

Hypothesis: That the low detail geometry will be more stiff as there are less polygons to allow the subtle folds of cloth to move freely.

3. Stiffness Forced Choice Experiment

Aim: To capture people's judgement of the actual material properties of the cloth for each LOD type.

4. LOD Comparison Experiment

Aim: To directly compare each of the representations in order to determine which of them most resembled the optimal simulation, which is important for a hybrid system that simultaneously displays lower and higher levels of detail. Hypothesis: That the impostor will resemble the high resolution more.

5. Impostor Update Frequency Test

Aim: To find the appropriate sampling rate (i.e., the number of viewpoints) for the impostors. Hypothesis: Characters with protruding limbs would need more viewpoint updates, as well as characters with very soft cloth.

6. System Experiment

Aim: To evaluate the effectiveness of the results by placing the characters in crowds of different sizes. Hypothesis: That the bigger the crowd, the more unlikely users would be to perceive differences in the LOD.

Apparatus, Setup and Stimuli Creation

The experiments were performed on two different machines. When the experiment involved the playing of movies a commodity PC was used with a 17 inch LCD monitor placed alongside a 24 inch widescreen LCD monitor to display the reference movies at the same time, as shown in Figure 2. We manually recorded the user input in these experiments. The experiments which used the OpenGL test application were displayed on a 21 inch CRT monitor, attached to a high end PC with a high performance graphics card. User input to this application was captured using a USB gamepad with right and left trigger buttons that allowed participants to make their selections.

We used commercial software [CloFX] to obtain our cloth simulations, but any high quality offline simulator could be used to produce the cloth animation. The rendering of the human's clothed mesh and impostor representation is based on the approach used by Dobbyn et al. [DHOO05]. We pre-simulate the deformation of both the virtual human's skin mesh using linear blend skinning and its cloth mesh us-

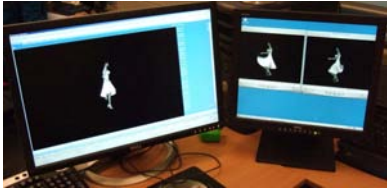


Figure 2: Setup for Stiffness Forced Choice experiment

ing ClothFX, based on the motion of its underlying skeleton. These meshes are then exported and stored in separate keyframed meshes or ‘poses’. At run-time in our experiment application, the corresponding pose is displayed depending on the virtual human’s current frame of animation. By pre-calculating and storing the deformation of the skin and cloth mesh in poses, we avoid the cost of deforming the meshes at run-time. In the case of displaying animated cloth using low resolution geometry, the poses are generated by simulating the cloth with a less detailed mesh. While this produces a slightly different animation to the high resolution cloth, this was the only feasible solution, as otherwise simplifying each pose of the high resolution poses resulted in rendering artifacts caused by topological differences between each pose’s triangles.

Generating the impostor representation involves capturing two types of images from a number of viewpoints around the model: a detail map image to capture the detail of the model’s diffuse texture and a normal map image whereby the model’s surface normals are encoded as RGB values. Since we only wish to display the humans at camera height (at least for the experiment), we need only generate images across a 1-dimensional slice of the full 2-dimensional space of viewing angles, as explained later (Figure 6). These images are then packed into a single texture for each keyframe of the animation. At run-time, the corresponding viewpoint for both types of images is selected depending on the virtual human’s current frame of animation and its position with respect to the camera. The selected images are texture-mapped on to a quadrilateral dynamically orientated towards the camera and are used to dynamically shade the impostor representation using programmable graphics hardware. A similar technique is detailed in [DHOO05]. The animation we chose was jumping on one foot as this displayed how the cloth deformed with air as well as showing its response when colliding with the character’s body. We chose a short loose skirt because it exhibited folds in the material effectively (see Figure 1).

3.1. Determining Perceptually Linear “Stiffness” Scale

ClothFX allows the user to change different properties of the garment in order to simulate different material types. We wanted to find material parameters that showed a good variety of different effects on the deformation of the skirt. With much trial and error we identified a soft material that could be turned into a very stiff material by altering just two material parameters: a u-v bending property that allowed greater resistance to bending, and a stiffness parameter. We then in-

terpolated the properties between the softest and stiffest materials to create 10 equally spaced steps.

We simulated the 10 different stiffness levels of material on the same skirt, with the same underlying motion, and rendered them in 3D Studio MAX as 3-second movies. As a pilot study we placed the 10 movies randomly on the screen, each playing in a loop. Six graphics experts were asked to put the movies in order, with the softest skirt on the left, and the stiffest skirt on the right of the screen, and place each of the intermediate steps in order between them. We will refer to the least stiff cloth as a soft cloth with a stiffness level of 1 while the stiffest cloth’s level is 10. Based on the results we found some clusters of similarity in the ordering. Stiffness levels 4, 5 and 6 seemed to form one cluster as they were placed together but in different orders, steps 7 and 8 formed another cluster, and finally steps 9 and 10 seemed to cluster. There was a clear distinction between these clusters, suggesting that the clustered materials were too similar to distinguish robustly. The 3 softest materials were always placed in the correct order, indicating that the steps at this level were more easily distinguishable.

From these preliminary results, we were able to develop a more robust perceived stiffness scale. Instead of evenly spacing the parameter intervals, we spaced them based on the preliminary results. We increased the number of samples nearer to the soft material end, as people were able to distinguish those materials more easily, and decreased the number of samples in the clusters that the participants had trouble in ordering, resulting in a new set of 8 steps (Figure 1). We performed the above test again with three new graphics experts, with the 8 3D Studio MAX rendered movies on screen at one time, and participants found the task relatively difficult but managed to order the steps correctly or with only one error in under two minutes. We then felt that we had a suitable stimulus set to form the basis for our various discrimination tasks, so we used them to export the different Levels of detail.

The 3D Studio MAX rendered cloth that was used previously will be considered as our *gold standard*. The gold-standard movies represent the ideal - they were rendered in 3D Studio Max with non real-time rendering and lighting. The high resolution model used in our rendering system is the same model, using the same number of polygons as that shown in the gold-standard movies. However, it was lit in our system using OpenGL lighting, and exported at discrete poses that were imported into our rendering system as keyframes. This is common practice for real-time characters in games and crowds, as it balances simulation speed and memory footprint. We will refer to the exported high resolution geometry as the *high resolution* representation. The 8 high resolution meshes of 8983 polygons each (6172 for the human model and 2811 for the skirt) were exported into a real-time animation system, and the corresponding 8 sets of *impostor* animations were generated.

Using the same jumping motion and with the same human model we regenerated the 8 different material animations,

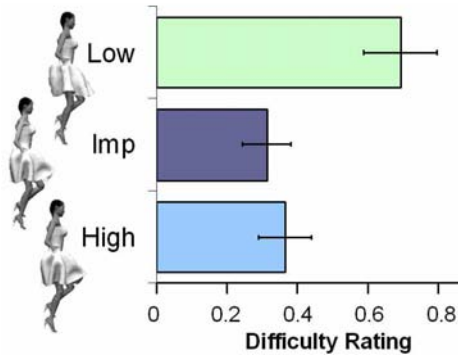


Figure 3: Results of stiffness sorting experiment

but this time we used a *low resolution* skirt of 509 polygons as the cloth. We then created 24 3-second movies of these animations from our rendering system, all at the same distance from the camera, from the same viewpoint and using the same lighting. All models were rendered in greyscale to eliminate any bias due to color, and with a black background to provide good contrast.

3.2. Stiffness Sorting Experiment

The second experiment aimed to establish whether the real-time lower detail cloth representations could produce the same range of distinguishable levels of stiffness as the gold standard model rendered offline. Twenty-four participants took part in this study (6 females, 18 males, aged between 17 and 27). They included undergraduate students, graduate students and professionals from different educational backgrounds. None of our participants were from the graphics lab and all were uninformed as to the purpose of the experiments. Most had an interest in but were not experts in computer graphics.

A between-groups design was used for this experiment, where each participant viewed either impostor, high detail geometry or low detail geometry. Eight movies, all showing one of the representations, were placed randomly on a 21-inch widescreen monitor, each playing in a loop. Similar to Experiment 1, participants were asked to place the 8 movies in order of stiffness, with the least stiff on the left and the most stiff on the right. The order in which the participants placed the movies was recorded and compared to the actual sequence.

To analyze these results we assigned a penalty of x for each of the movies that were not placed in the correct order, with x corresponding to the number of places that the movie deviated from its correct placing in our perceived stiffness scale. A two-factor ANOVA [How99] was performed over the full dataset to get an overall impression of whether there was a difference between how the participants sorted the different cloth LOD representations. Figure 3 shows us that overall the participants found the low detail geometry cloth more difficult to sort than the high detail ($F_{1,112} = 5.65, P < 0.02$), and the impostor representation ($F_{1,112} = 8.69, P < 0.004$). There was no statistical sig-

Stiff Level	LOD Comparison	Statistical Significance
Softest	Impostor < Low	$F_{1,46} = 3.15, P < 0.08$
Softest	High < Low	$F_{1,46} = 4.16, P < 0.05$
2	Impostor < Low	$F_{1,46} = 22.4, P \approx 0$
2	High < Low	$F_{1,46} = 22.4, P \approx 0$
3	Impostor < Low	$F_{1,41} = 20, P \approx 0$
3	High < Low	$F_{1,46} = 20, P \approx 0$
4	High > Impostor	$F_{1,46} = 4.46, P < 0.05$
4	Impostor < Low	$F_{1,46} = 19.6, P \approx 0$
4	Low > High	$F_{1,46} = 5, P < 0.05$
Middle	High > Impostor	$F_{1,46} = 26, P \approx 0$
Middle	Impostor < Low	$F_{1,46} = 47, P \approx 0$
7	High > Impostor	$F_{1,41} = 3.5, P < 0.07$
Stiffest	Impostor < Low	$F_{1,46} = 5.3, P < 0.05$

Table 1: Statistically significant ANOVA results for stiffness forced choice experiment. (All other pairs are insignificant)

nificance between sorting the impostor and sorting the high resolution movies.

These results indicate that the perception of subtle differences in cloth motions using the impostor is closer to that of the high detail geometry cloth simulation than the low geometry simulation. This also supports the findings in [MDO05] that impostors are better at depicting small differences in human motion.

3.3. Stiffness Forced Choice Experiment

The third experiment we performed was a test of how well the high and low detail geometry and impostor reproduced the stiffness levels of the gold standard cloth. Participants in this experiment were the same as those in Experiment 3. They were shown 2 gold standard movies (that of the stiff and that of the soft skirts) beside each other on one 17 inch monitor (Figure 2). A slightly different jumping animation was used in our gold standard animations and they were captured from a different viewing angle, as we felt that participants might otherwise compare exact silhouettes of the cloth rather than getting a feel for the material properties of the cloth. We created four playlists with three repetitions of each of the eight cloth stiffness levels on each of the three LOD representations (high, low and impostor) randomly interleaved, with a three second blank video in-between each trial. A total of (7, 6, 6, 5) participants saw playlists 1-4 respectively. We used a Two Alternate Forced Choice (2AFC) design for this experiment. Participants were asked after every trial to indicate which of the two gold standard animations (stiff or soft) was more similar to the current target animation, playing on an adjacent monitor.

The experimenter sat beside the participant and manually recorded their choices. As there were only 4 different playlists, we began our analysis by testing if there were any ordering effects due to the playlist. We found no statistical significance due to the effect of playlist order with the high and low resolution cloths. However, we did find a significant effect for the impostors ($F_{3,20} = 2.64, P = 0.08$). On closer analysis of the effect of playlist order on the impostor we found no statistical significance for playlists 1-3, however

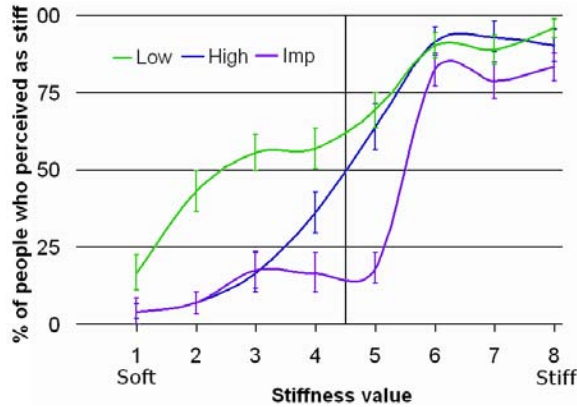


Figure 4: Perceived stiffness for different LOD cloths. Standard error bars are shown

we found that playlist 4 did show ordering effects for impostors at stiffness level 3, 6 and 7. We simply removed the five values for each of these stiffness levels during our comparison. This is the reason for the degrees of freedom differences for the impostors in Table 1. Once this data was removed, we plotted the percentage of stiffness choices for each stiffness level for each LOD representation (Figure 4).

We tested for a LOD/Stiffness interaction effect and found ($F_{14,432} = 4.09, P \approx 0$). There was also a significant overall LOD effect ($F_{2,432} = 37, P \approx 0$), and an overall stiffness effect ($F_{7,432} = 87, P \approx 0$). We then performed pairwise single factor ANOVAs between each LOD pair at each stiffness level, and the statistically significant differences are reported in Table 1.

The interpretation of these results is evident from Figure 4. Participants found that the perceived stiffness of the cloth motion for the impostor was closer to that of the high resolution than the low resolution for low stiffness levels (i.e., soft materials). This suggests that the impostor better matches the high detail geometry motion at low stiffness levels. There is a divergence at the middle stiffness levels, where most participants rated the impostors to be soft, and few found the low detail to be soft. At the high stiffness levels, more participants' perception of the low resolution cloth motion was closer to the high detail geometry than the impostor.

Overall, there seems to be a tendency for the participants to perceive the low detail geometry cloth motion to be stiffer than the high geometry, and the impostor to be less stiff than the high geometry. The impostor curve appears to be displaying a step function, where participants were certain of their perception at the lower stiffness levels and also the higher levels, with no intermediate level. The high resolution's curve displays a smoother graduation of certainty. The low resolution curve is more difficult to interpret, but it is clear that participants found the cloth motions stiffer overall, as we initially hypothesized. The fact that participants perceived the impostor cloth motion as being less stiff than the high resolution motions is a surprising result. However, this

may explain why McDonnell et al. [MDO05] found that participants perceived small arm motion differences more easily when viewing the impostor than in the high resolution mesh.

3.4. LOD Comparison Experiment

The previous experiments gave us a good idea about how well different levels of detail reproduced stiffness levels, but they did not show us which best emulated the ideal - crucial in a hybrid system where different LOD human representations are being displayed. Therefore, a third experiment was performed in order to see how well each representation matched the gold standard.

The participants that took part in the previous experiments also took part in this experiment. An OpenGL test application was used for this experiment, where participants were asked to compare pairs of cloths rendered at the three different resolutions; high resolution, low resolution or impostor. Hence, they were able to directly compare each of the representations with each other, which gave us a good idea of how different they appear to the viewer. They were first shown pairs of the most rigid cloth and were asked which cloth animation was most similar to the gold standard rigid cloth. The animation of the gold standard used a slightly different motion and was shown from a different viewpoint and played on an adjacent screen. Participants pressed left or right on the gamepad to choose the most similar simulation. Three repetitions of the pairs were played at a distance that was equal to the 1 to 1 pixel to texel ratio of the impostor (i.e., the point at which Hamill et al. [HMDO05] found that their impostors were perceptually equivalent to a high resolution polygon model), thus they were not basing the comparison on artifacts due to the impostor rendering. It should be noted that the impostor and the high resolution model are not exactly the same as the impostor's normals are encoded as RGB values, and are therefore approximations of the normals of the high resolution polygon model. Participants were then shown pairs of the cloth with stiffness approximately halfway on our estimated scale (i.e., the fifth image in Figure 1), and were asked to compare them with the corresponding gold standard cloth. Four repetitions were recorded, and pairs were shown at random and on random sides of the screen. Finally, the participants viewed pairs of the most soft cloth and were asked to compare them to the gold standard as before.

The number of times that a participant preferred each LOD representation over each of the others was recorded. A 2-factor ANOVA was first performed on the results for the forced choice comparison of the high resolution cloth to the low resolution to get an overall idea for how often one representation was chosen over the other.

An LOD/Stiffness interaction effect was found ($F_{4,216} = 2.94, P < 0.03$), and it was found that participants chose the high as being more similar to the gold standard than the low overall ($F_{2,216} = 137.78, P \approx 0$). We then performed pairwise single factor ANOVAs for further insight. Participants chose the high resolution as being more similar to the gold

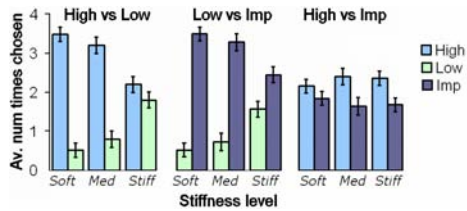


Figure 5: Results for LOD Comparison Experiment

standard than the low for the most soft ($F_{1,48} = 129.87, P \approx 0$) and the middle material ($F_{1,48} = 66.46, P \approx 0$), however there was no statistical significance when they compared high to low when viewing the stiff material (Figure 5). If we compare these results to those of the previous experiment, shown in Figure 4, we can see that it was in fact the case that participants found the low resolution cloth different from the high resolution for the very soft and middle materials, but that they were very similar when the material was stiff. This is to be expected because of the degrees of freedom reduction in low resolution cloth, resulting in a loss in deformation of the soft material.

A 2-factor ANOVA was then performed on the participants' dataset for the comparison of the high resolution cloth motion to the impostor. It was found that, overall, participants chose the high resolution as being more similar to the gold standard than the impostor ($F_{1,144} = 13.9, P < 0.0003$). Pairwise single factor ANOVAs showed that participants were equally likely to choose the impostor as the high resolution for the softest material. However, for the middle ($F_{1,48} = 6.22, P < 0.02$) and stiffest materials ($F_{1,48} = 7.08, P < 0.02$), they chose the high resolution cloth more often. This trend to prefer the high resolution mesh for stiffer and middle materials is supported by the fact that participants in the stiffness forced choice experiment (Figure 4) found these stiffness levels significantly different.

Finally, we analyzed the comparison of the low to impostor, and found that participants chose the impostor as more like the gold standard for the softest material ($F_{1,48} = 129.87, P \approx 0$), middle material ($F_{1,48} = 67.7, P \approx 0$), the stiffest material ($F_{1,48} = 9.62, P < 0.004$) and overall ($F_{1,144} = 167.32, P \approx 0$). Again, this is supported by the fact that participants in the previous study found these stiffness levels statistically different.

Our conclusion from these results is that, when viewing these representations in a hybrid system that simultaneously displays virtual humans using two types of representation, switching intermittently between them, the low resolution will not resemble the high resolution as closely as the impostor does, thus resulting in significant artifacts. In Section 3.6, we will validate this finding with crowds of humans of varying scales.

3.5. Impostor Update Frequency Test

From the results recorded, we can see that impostors are good at representing the deforming folds of cloth and are a good substitute for high resolution geometry clothed mod-



Figure 6: Generating impostor images from a camera positioned on the circumference of a circle every 45° .

els at a 1:1 pixel to texel ratio distance from the camera. As mentioned above, impostors are generated by rendering multiple images of the human from different viewpoints for every frame of animation. The appropriate viewpoint is selected with respect to the camera in the real-time system. Typically, these viewpoints are generated at regular intervals around a sphere, so the sampling density can be described by the number of degrees difference between each segment of the sampled sphere (e.g., Figure 6 shows impostor images generated every 45°).

Ideally, impostors would be generated at very small intervals around a sphere, the same number of times that a polygonal model would be updated, which would allow seamless transitions between the images. However, as we are using pre-generated textures, texture memory consumption prevents choosing such a dense sampling, so there is a need to pick an optimal number of viewpoint images to generate. Dobbyn et al. [DH005] and Tecchia et al. [TLC02] report rendering 17 and 16 viewpoints of their impostors from one side of the human and mirroring the impostors for the reverse angle. This corresponds to an update rate of 10.58° (180° divided by 17) and 11.25° respectively. However, they do not specify their reasons for choosing these numbers of viewpoints. With the addition of cloth to the impostors, mirroring is no longer possible due to the non-symmetric nature of cloth (the cloth on one side is usually not identical to the other side, due to the folds occurring in different places). Also, it was not clear that this directional sampling density would be appropriate when clothing was added to the impostors. As in Tecchia et al. and Dobbyn et al., interpolation was not used between different views, as it would be computationally intensive and could introduce visual artifacts.

Hamill et al. [HMDO05] performed psychophysical experiments to find the sensitivity of participants to the update rate of dynamic building impostors while orbiting them. They reported that a sampling density of 8° was the point at which people were 50% likely to find the rotation smooth, which was found to be the case for all of the buildings tested. We conducted a set of psychophysical experiments similar to theirs in order to determine the threshold of acceptability for viewpoint updates of human clothed impostors. All participants that took part in the previous experiments also took

part in this experiment along with a further 24 (4 females, 20 males). The experimental stimuli were presented on an OpenGL test application. A white background was used, and the models were presented on a grid as a ground plane.

Stimuli: We chose 6 characters as stimuli for this experiment (Figure 7): the first two models were the jumping character used in the previous experiments, with the soft and stiff skirts, and were chosen so that we could plug the results back into the system, and also to test the effect of the 2 cloth stiffness levels on the perception of viewpoint changes. We will refer to these as *Jumping Soft* and *Jumping Stiff*. The next two characters used the same model, but with a long skirt, and with a walking motion rather than a jumping one. The two test cases were: arms firmly by her side, and arms stretched out in front of her. These two cases were chosen to test the effect of different dimensions on the same character. We will refer to these as *Arms Down* and *Arms In*. The fifth character, *Skeleton*, had a long flowing cloak, this character was very long and thin, and we hypothesized that this character would need more updates as its dimensions were irregular. This character’s motion was a flying motion, with both arms moving slowly up and down. The final character we tested was a female model, *Edith*. With a long tight dress and normal walk, this character was used as it is typical of the type of character used in a pedestrian crowd scene.

We pre-generated 256 impostors for every frame of the 10-frame animation, which corresponded to an *optimal* sampling density of 1.4° (i.e., 360° divided by 256). The characters moved on a circular path at a normal walk pace, with the closest point to the viewer on the circle being the pixel to texel ratio distance reported in [MDO05], as impostors would not be viewed any closer than this in a crowd system. The character was placed at a random point on this path, and walked for 3 seconds, in either a clockwise or an anticlockwise direction. Participants were asked to specify, using the gamepad, whether they perceived the change in orientation of the character as jerky or smooth. A screen with “Smooth” written on the left, and “Jerky” on the right popped up in between each trial, to prompt the participant to press the corresponding left or right trigger on the gamepad. The next trial they saw depended on their previous response.

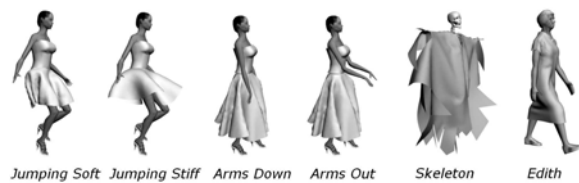


Figure 7: Six different characters used in impostor update frequency test

Procedure: A staircase procedure is often used for gathering experimental data efficiently in a psychophysical experiment [Cor62, Tre95]. A stimulus is alternately increased

or decreased until it passes a participant’s detection threshold.

We randomly interleaved an ascending and descending staircase for each model. The ascending staircase began with an update frequency of 1.4° and the descending 45° , i.e., every time the orientation had changed by this amount, a new image was shown. An up-down staircase was employed so that, for every correct response, the stepsize was added to the current step, and for every incorrect response, the stepsize was subtracted from the current step. A response was considered correct if they chose smooth, and incorrect if they chose jerky. The initial stepsize was 11.25° , and was successively reduced to our lowest sampling step of 1.4° by halving at the first 4 reversals. A reversal occurred when a participant’s previous response was different to the current response. Thus, the stepsize was refined when closer to the area of interest. Eight more reversals were then counted before the experiment ended.

For each participant, a psychometric curve was fitted to their dataset as described in [HMDO05]. This allowed us to find the Point of Subjective Equality (PSE).

	Skel.	ArmsOut	ArmsDown	JumpSoft	JumpStiff	Edith
μ	12.2	18.8	21.6	18.9	20.9	26.8
σ	3.9	10.4	6.8	12.1	12.4	11.9

Table 2: Mean PSEs and standard deviations of all models

$$f(\theta; \mu, \sigma) = \frac{1}{2\pi\sigma} e^{-\left(\frac{\theta-\mu}{2\sigma}\right)^2} \quad (1)$$

The PSE is the point at which participants were equally likely to find an animation smooth or jerky, i.e., where they have a 50% chance of considering that motion smooth. This value is found on the curve as the stimulus level (in this case the number of degrees) at the 50% detection level. A psychometric probability curve for each of the characters, derived from all of the data was then created, using the average of all participants’ PSE and standard deviations (Figure 8 and Equation 1). The mean is the estimated PSE (value with 50% likelihood of detection), the standard deviation allows other measures to be determined (e.g., the 80% level of acceptance). Each model’s μ and σ values can be seen in Table 2. These curves are very useful, as θ in Equation 1 can be replaced by any number of degrees in order to find the percentage probability of perceptual acceptance of that sampling density. We use this curve in Section 3.6 to find the 80% probability of acceptance for the impostors used in that experiment.

Analysis: Mean PSE values across all of the participants are illustrated in Figure 9. We observe that the mean PSE for *Skeleton* was the least, which implied that the participants needed many viewpoint changes in order to be able to tell the difference between smooth and jerky animations for this character. The mean PSE for *Edith* was the most, meaning that the fewest updates were needed for her. All of the other

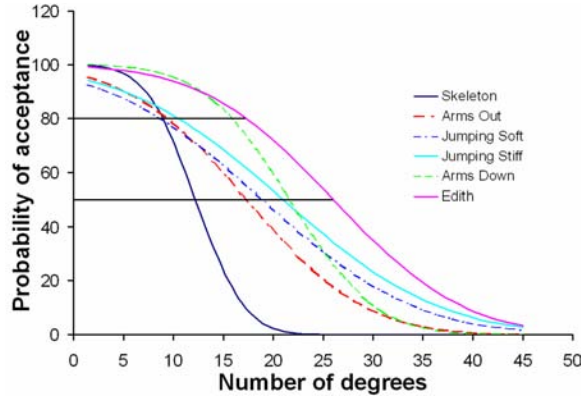


Figure 8: Probability of acceptance curve derived from psychometric data

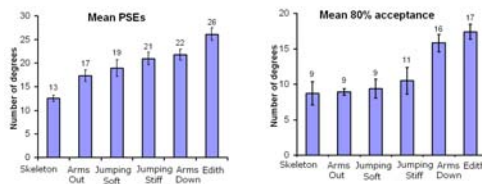


Figure 9: (a) Mean PSE values for all models, (b) Mean values for 80% probability of acceptance

character’s PSE values formed a group together in between the mean PSE of *Skeleton* and *Edith*.

While the mean PSE values give us insight into the point at which participants were equally likely to find the animation smooth as jerky, they do not tell us about the ideal number of viewpoint updates necessary to assist developers in choosing an acceptable update rate. We chose the 80% level as the most appropriate level to balance texture memory consumption and quality of animation (choosing higher levels would require impractical updates for a real-time system). Figure 9b illustrates the differences in the mean values at this level, for each of the model types. An overall ANOVA showed that there was an overall effect of model type on the 80% perception level of smoothness ($F_{5,121} = 9.13, P \approx 0$). It is clear from this graph that the characters are clustered into 2 distinct groups at this threshold. This is confirmed by ANOVA comparisons which showed that there was no statistical significance between the perception of smoothness for the normal walking characters *Arms Down* and *Edith*. Also, there was no significant difference between the other four characters. However, there was a significant difference between the walking characters and all of the others together ($F_{1,125} = 44.6, P \approx 0$). This clear difference between the 2 groups at this threshold is very interesting. *Arms Down* and *Edith* both had a normal walk, but had different clothing. *Edith*’s clothing was almost like that of a skinned character, as it displayed little secondary motion and was tight fitting. *Arms Down*’s clothing was much looser, however, suggesting that clothing stiffness did not have an effect on the per-

ception of smoothness. This is also supported by the fact that there was no difference between the perception of smoothness when viewing *Jumping Soft* and *Jumping Stiff*.

This suggested that it was not the clothing that affected the perception of smoothness, but the width to depth ratio of the characters, as *Jumping Soft*, *Jumping Stiff*, *Skeleton* and *Arms Out* had either protruding or outstretched limbs, making their width to depth ratios higher. Averaging the values for each of the groups gives us the final viewpoints necessary for developers.

Developer Guidelines: For normal walking characters, with either stiff or soft clothing, a viewpoint update rate of 17° is necessary to guarantee with 80% likelihood that users will not notice viewpoint changes of the impostors. This corresponds to 21 images that need to be generated at equal spacing around the character. We suggest rounding to the nearest even number of images (22) in order to include the direct front and the direct back images, particularly in applications where a front-on view would be most noticeable. For other characters whose width to depth ratios are large, a viewpoint update rate of 9° is advised. This corresponds to 40 images around the character. In [DHO05] and [TLC02], updates of 10.58° and 11.25° were used. We can now see that these rates were underestimates for complex characters but overestimates for normal walking characters.

3.6. System Experiment

The final experiment was a system evaluation, to test the validity of the results of the previous experiments in a real crowd scenario.

The crowd scene was a typical town square scene, with an open area for the characters to be placed, and buildings surrounding them. The scene was populated with the female jumping model used in Experiments 1-5. Each model was randomly placed in the scene, at a random orientation, and jumping on the spot. The characters in the scene were either all wearing the most stiff skirt from the experiments, or the most soft skirt. Each of the models were colored differently, as we wanted to test the results on a real scenario, where models and scenery are almost always shown in color.

The experiment included three typical crowd systems: full geometry, hybrid high polygon/impostor and hybrid high polygon/low polygon. In the full geometry crowd system, all characters were high resolution polygonal models of 8983 polygons each (6172 for the human model and 2811 for the skirt). The hybrid high polygon/impostor system contained the high resolution polygon models nearest to the camera, and the impostor representations at the back (Figure 10c). The latter were displayed at the pixel-to-texel ratio where they are perceptually equivalent to high resolution meshes [HMDO5]. This pixel-to-texel ratio was found for individual characters, but was never tested on a large crowd. We hypothesize that this ratio will hold true for crowds, as it could be considered a conservative estimate in this more complex scenario. We used the results of the previous experiment to choose the number of viewpoint images necessary for the two models.

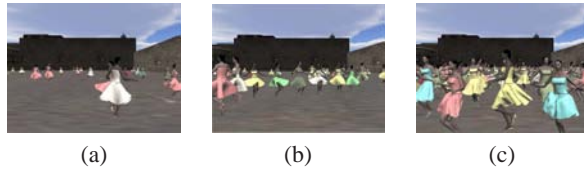


Figure 10: (a) Small crowd of *All Hi* with stiff cloth. (b) Medium crowd of *Hi/Lo* with stiff cloth. (c) Large crowd of *Hi/Imp* with soft cloth.

The hybrid high/low resolution polygon system contained high resolution characters at the front, and low resolution at the back (Figure 10b). We kept the female model the same as the high resolution model but used a lower resolution cloth, like in the previous experiments. This decision was made as it was the perception of cloth level of detail that we were interested in examining, and not that of the entire model. Five hundred and thirty polygons were chosen for the low resolution skirt. We based this on the findings of [MDO05], where a lower resolution mesh of 27.5% of the number of vertices of the mesh that created it was found to be perceptually equivalent to that high detail mesh, at the impostor pixel to texel ratio, which is the closest distance that we will be placing the lower detail representations.

In a typical hybrid crowd system, the LOD choice depends on the distance of a character from the camera. As the camera moves through the scene, switching between representations will occur, due to the camera distance changing. To examine this effect in our experiments, the camera zoomed up and down through a corridor between the characters at a speed of 4m/s, in order that LOD switching occurred. Switching between impostor viewpoints also occurred in this case.

The effect of switching between impostor viewpoints was then examined independently by allowing the camera to only pan from left to right at a speed of 2m/s, where the impostor distance was fixed. In this case, the impostor viewpoints were changing but no switching back and forth to high resolution geometry occurred.

We wanted the framerate to stay constant throughout the experiment, in order that the camera speed did not bias the results, so we created movies of the system rather than letting the participant view the actual system. One hundred and eight 4-second movies were created in total:

3 types of system: *All Hi*, *Hi/Lo*, *Hi/Imp* \times 2 skirt types: most stiff and least stiff \times 3 crowd sizes: small (50 humans), medium (100 humans) and large (1000 humans) \times 2 conditions: camera panning from left to right, camera moving up and down a fixed corridor \times 3 random placings: 3 different random placings of characters in the scene.

The 108 different movies were randomly placed in a playlist, with a 1-second blank screen in-between. A different playlist was generated for each participant. Thirteen participants took part in this experiment (2 females, 11 males, aged between 17 and 28). Most had an interest in but were not experts in computer graphics, and were from different educational backgrounds. Each participant viewed the se-

quence, and was asked for every trial, whether all of the characters in the scene were the same or if they noticed that some of the characters looked different in any way. We did not specify what exactly was meant by “different”, as we felt that it would be better to get the participant’s unbiased view of the system, rather than giving them something to look out for. The experimenter sat beside the participant and recorded their responses to the stimuli. A value of 1 was recorded for “different” and 0 for “same”.

LOD & Camera

We first analyzed the effect of camera panning or zooming on LOD. We hypothesized that participants would be more aware of the differences in representations in the hybrid systems when the camera was zooming, as switching between representations would highlight the differences. A 2-factor ANOVA was performed on the dataset and it was found that there was no significant overall effect of camera motion on the ability of participants to tell the differences between the representations. This supported the null hypothesis, and implied that participants were unaware of the switching between representations, which is very good news for hybrid systems. However, there was a significant effect of LOD ($F_{2,1398} = 332, P \approx 0$).

LOD & Stiffness

Stiffness had an interaction effect with LOD in our previous experiments for individual characters, so we hypothesized that this would also be the case for crowds. In particular, we felt that the low detail geometry would be noticed fewer times for the stiff skirt than for the soft skirt. After performing a 2-factor ANOVA, an interaction effect was found ($F_{2,1398} = 17.6, P \approx 0$). There also was an effect of LOD ($F_{2,1398} = 338, P \approx 0$), but no effect of stiffness (Figure 11).

For the stiff cloth, participants noticed the difference between the hybrid *Hi/Lo* and *All Hi* movies significantly more times ($F_{1,466} = 109, P \approx 0$). Also, they noticed the difference between *Hi/Lo* more times than the *Hi/Imp* ($F_{1,466} = 115, P \approx 0$). There was no statistical significance between the number of times that they noticed a difference in *All Hi* compared to *Hi/Imp*.

For the soft cloth, there was a difference between *All Hi* and *Hi/Lo* ($F_{1,466} = 427, P \approx 0$), and between *Hi/Lo* and *Hi/Imp* ($F_{1,466} = 373, P \approx 0$). Again, there was no difference between *All Hi* and *Hi/Imp*.

As expected, there was a statistically significant difference between *Hi/Lo* stiff and soft, with the soft cloth low geometry being noticed more times than the stiff cloth. There was also a difference between impostor stiff and soft - with differences in the soft being noticed fewer times than differences in the stiff cloth. Similar differences in stiffness were present for *All Hi*.

LOD & Crowd Size

Experiments 1-5 all depicted scenes with only 1 or 2 characters. This represents the worst-case scenario, as the character was being analyzed directly, with no surrounding distractions. Two hypotheses were considered: the first was that an increase in crowd size would highlight the difference in

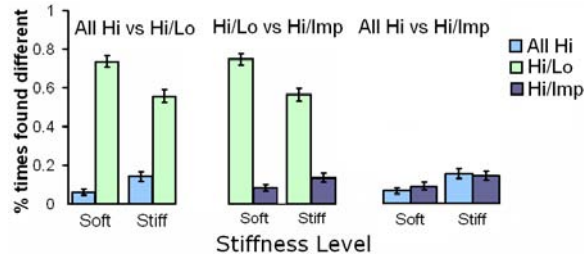


Figure 11: System Experiment: LOD vs Stiffness

model representation in the hybrid systems, as the number of these models would be greater, and possibly make them more noticeable. The other possibility considered was that the increase in crowd size would mask the differences in the models, as there would be more high resolution geometry at the front, and the large crowds might distract attention away from the individual characters. Our results showed that neither was the case. Surprisingly, it was found that there was no overall effect of crowd size, nor was there any interaction effect of crowd size with LOD. Again, there was an overall effect of LOD ($F_{2,1395} = 330, P \approx 0$).

Stiffness & Crowd Size

Finally, the effect of stiffness and crowd size was tested. We hypothesized, based on the previous analysis, that there would be no significant effect of either, and we were correct.

4. Conclusions and Future Work

We have presented a series of psychophysical experiments on LOD cloth. The sorting experiment was needed to determine if each representation in our system could reproduce distinguishable stiffness levels, as the gold standard could. Results showed that overall, the impostor and high resolution reproduced the stiffness levels effectively, whereas the low resolution was less adequate. The Stiffness Forced Choice experiment aimed to capture people's judgements of the actual material properties of the cloth for each LOD type. Results showed that overall, participants tended to consider the low resolution to be stiffer than the high resolution, and the impostor to be less stiff. The Direct Comparison experiment was performed in order to determine which of the levels of detail most resembled the ideal solution - crucial in a hybrid system where different LOD human representations are being displayed simultaneously. It was found that, when comparing the impostor to the low resolution, the impostor was always closer to the ideal, and therefore the most appropriate candidate to replace the high resolution (i.e., the best real-time representation for characters close to the viewer), when needed.

After establishing these facts, we found a metric for the most appropriate update frequency for different impostor types. Our impostor update frequency experiment gave us a perceptual metric for choosing the optimal number of viewpoint images rendered at ground height. As a result of this research, game developers and other practitioners will now be able to generate impostors with minimal jerkiness between viewpoint images. Our results were validated in the Sys-

tem Experiment, as participants' perception of differences in characters was the same for the hybrid high/impostor system as for the system with all high resolution geometry (i.e., they did not notice the viewpoint changes). While our evaluation was carried out on a pre-generated impostor representation, the same metric could be utilized in deciding when a dynamically generated impostor needs to be updated. As we only tested for one elevation, we would also like to find the optimal sampling rate for different camera elevations.

Finally, we tested the effectiveness of the different representations in a crowd system, with different crowd sizes. The results of these experiments complement the results for a single character of the LOD Comparison Experiment, as it was found that participants noticed the differences between low detail and high detail geometry more than impostors and high detail geometry. Furthermore, the results of the Stiffness Forced Choice Experiment were confirmed, as it was seen that low detail geometry for stiff cloth is less noticeably different from high detail cloth than soft cloth. We also found a surprising result, that crowd size did not effect the perception of differences between characters in hybrid systems.

References

- [BTH*03] BHAT K. S., TWIGG C. D., HODGINS J. K., KHOSLA P. K., POPOVIC Z., SEITZ S. M.: Estimating cloth simulation parameters from video. *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on computer animation* (2003), 37–51.
- [CloFX] Cloth simulation software, ClothFX.
- [Cor62] CORNSWEET T.: The staircase method in psychophysics. *American Journal of Psychology* 75 (1962), 485–491.
- [dHCUCT04] DE HERAS CIECHOMSKI P., ULICNY B., CETRE R., THALMANN D.: A case study of a virtual audience in a reconstruction of an ancient roman odeon in aphrodisias. *The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heiritage (VAST)* (2004), 9–17.
- [DHOO05] DOBBYN S., HAMILL J., O'CONOR K., O'SULLIVAN C.: Geopostors: a real-time geometry / impostor crowd rendering system. *SI3D '05: Proceedings of the 2005 symposium on Interactive 3D graphics and games* (2005), 95–102.
- [HMDO05] HAMILL J., McDONNELL R., DOBBYN S., O'SULLIVAN C.: Perceptual evaluation of impostor representations for virtual humans and buildings. *Computer Graphics Forum* 24, 3 (2005), 623–633.
- [How99] HOWELL D.: *Fundamental Statistics for the Behavioral Sciences*. Duxbury Press, 1999.
- [MDO05] McDONNELL R., DOBBYN S., O'SULLIVAN C.: LOD human representations: A comparative study. *Proceedings of the First International Workshop on Crowd Simulation* (2005), 101–115.
- [TLC02] TECCHIA F., LOSCOS C., CHRYSANTHOU Y.: Visualizing crowds in real-time. *Computer Graphics Forum* 21, 4 (2002).
- [Tre95] TREUTWEIN B.: Minireview: Adaptive psychophysical procedures. *Vision Res.* 35, 17 (1995), 2503 – 2522.