DOI 10.1002/cav.1728

RESEARCH ARTICLE

Perception of collisions between virtual characters

Sybren A. Stüvel | A. Frank van der Stappen | Arjan Egges*

Virtual Human Technology Lab, Universiteit Utrecht, Utrecht, The Netherlands

Correspondence
A. Frank van der Stappen , Virtual Human Technology Lab, Universiteit Utrecht, Utrecht, The Netherlands.
Email: a.f.vanderstappen@uu.nl

Abstract

With the growth in available computing power, we see increasingly crowded virtual environments. In densely crowded situations, collisions are likely to occur, and the choice in collision detection technique can impact the perceived realism of a real-time crowd. This paper presents an investigation into the accuracy of human observers with regard to the recognition of collisions between virtual characters. We show the result of two user studies, where participants classify scenarios as "colliding" or "not colliding"; a pilot study investigates the perception of static images, whereas the main study expands on this by employing animated videos. In the pilot experiment, we investigated the effect of two variables on the ability to recognize collisions: distance between the character meshes and visibility of the inter-character gap. In the main experiment, we investigate the angle between the character paths and the severity of the (near) collision. On average, respondents correctly classified 72% (static) and 68% (animated) of the scenarios. A notable result is that the maximum uncertainty in determining existence of collisions occurs when the characters are overlapping and that there is a significant bias towards answering "not colliding." We also discuss differences in bias in the recognition of upper- and lower-body collisions.

KEYWORDS
perception, collision detection, virtual humans, believability, performance

1 | INTRODUCTION

Animating a crowded scene, such as a busy shopping street, evacuation scenario, or large procession, requires tight packing of virtual characters. In such cases, collisions are likely to occur. The choice in collision detection technique can impact the maximum density, which can be handled in real-time, and perceived realism of the crowd.

Many collision detection schemes aim at exactness, which is vital in areas like computer-aided design and robotic product manufacturing. Such exactness may not be the best approach for collision detection between virtual characters. People observing virtual characters may not be able to recognize collisions in certain configurations, and thus specific optimizations could exploit this to improve collision detection performance without sacrificing perceived quality, or to provide a better match between observed and detected collisions. Exactness seems even less crucial in crowds of virtual characters; people observing a crowd of virtual characters do not always have all the information to determine whether a collision occurs.

1.1 | Main contribution

This article presents an investigation into the accuracy of human observers with regard to the recognition of collisions between virtual characters. We have performed two user studies into the perception of collisions between virtual characters, to determine how accurate human observers can classify a situation as "colliding" or "not colliding." A pilot experiment investigates the perception of static images. The main experiment uses video to explore the effects of movement; we have investigated the angle between the character paths and the severity of the (near) collision, and present a statistical model for the expected accuracy.

Our results show that the average observer has a bias towards negative ("not colliding") answers, mostly in cases of minor collisions, and that the accuracy of the answers has an asymmetrical relation with the severity of the (near) collisions. To conclude, we suggest a technique to improve performance of collision handling possible collision shape and a simplification scheme that matches human perception. This simplification is based on inner approximations rather than the coarse cylindrical outer approximations that are commonly used in animations.

Furthermore, we investigate the difference in perception of lower-body and upper-body collisions, which, to our knowledge, has not been explored yet.

1.2 | Organization

The remainder of this article is organized as follows. Section 2 discusses related work. The overall experiment design is described in Section 3. The pilot experiment is described in Section 4, and the main experiment in Section 5. The implications are discussed in Section 6. Section 7 concludes the paper.

2 | RELATED WORK

The perception of collisions between large numbers of objects was studied by O'Sullivan et al.[1,2] They showed that when the simulation becomes more complex, observers rely on their own naïve or common-sense judgements of dynamics, which are often inaccurate.[2] We will come back to this later in this section.

Perceptual studies of virtual characters have been performed with regards to motion, emotion, timing, and sound.[3–5] An experiment by Hoyet et al.[6] investigated the perception of causality in virtual interactions, dealing with pushing interactions between characters. Their focus lies on the perceived realism of a scene, after applying alterations commonly found in virtual environments such as games. In contrast, our experiment does not focus on perceived realism, but on whether collisions can be perceived at all.

Perception of collisions between a real user and a virtual entity has also been studied. DeLucia[7] investigated the perception of collision with respect to traffic safety, that is, collision between moving obstacles and a stationary observer. Olivier et al.[8] performed a user experiment to assess whether real humans are also able to accurately estimate a virtual human motion before collision avoidance and conclude that when an observer is in front of a simple display, judgement of crossing order was easier than recognition of future collisions. This shows that perceiving collisions, at least when one virtual character is involved, is nontrivial. They continue to show that the "bearing angle," the angle at which one entity sees the other, plays a large role in the perception of collisions.

Kulpa et al. present an experiment of both the perception of crowds of virtual humans, and an accompanying *level of detail* (LOD) technique for collision detection.[9] Their focus lies on the accuracy of human observers in recognizing collisions, based on various parameters such as camera distance, horizontal and vertical camera angle, and character distance. They measure the latter as the distance between the characters' root joints, which, although easy to compute, provides only a rough estimate for the distance between the two characters. In contrast, in this paper, we use the actual distance between the character shapes, as described by LaValle.[10] This metric also determines whether there is actually a collision or not. Another contrast to the aforementioned work by Kulpa et al. lies in the placement of the camera. We place the camera such that the collision itself is maximally visible. Furthermore, rather than having the characters walk along parallel paths, we consider crossing paths of the characters and measure the effect of the angle between those paths on the perception of the collision.

To speed up collision detection algorithms, it is common to forego the possibly complex shape of the object and use a simplified shape instead. LOD techniques can generate such shapes, most notably applied to model simplification for rendering acceleration.[11,12] LOD techniques have seen less emphasis in the area of collision detection and mostly focus on the simplification of the colliding shapes.[13–15] Otaduy and Lin[16] introduced a technique that also considers the velocity and view size of the objects, and allows for time-critical detection in a similar way as introduced by Hubbard.[17] Apart from the velocity-based LOD technique, these techniques do not focus on human perception of collisions, and taking this into account could lead to better algorithms. O'Sullivan et al.[18] incorporated LOD techniques not only in rendering and collision handling, but also in the animation and behavioral algorithms. In the discussion section, we explore possible adaptations of LOD techniques to bring them in line with our findings on human perception.

We have performed two user studies. The first experiment is a pilot experiment, the main goal of which is to determine reasonable ranges of parameters to be used in the main experiment. The pilot experiment uses static poses rather than animated characters, because it is nearly impossible to find animated cases corresponding to specific desired parameter values, even though we are aware of the limitation that the lack of movement may make it difficult to fully assess the situation. As such, the participants' answers may depend on common-sense judgements, which O'Sullivan and Dingliana[2] described as inaccurate. However, their conclusions are based on geometric shapes, and it will be interesting to see how accurate or inaccurate the results are for human shapes. The main user experiment uses animated characters, because we aim at applying our results to collision detection strategies for moving characters. We examine the effect of the severity of the (near) collision and the angle between the paths of the two characters.

3 | EXPERIMENT DESIGN

In this section, we describe the common experiment design, which is shared by both the pilot and main experiments. In both user studies, we show rendered 3D scenes involving two virtual characters in an otherwise empty virtual world. One of the characters is male, the other is female. The two characters are posed using previously recorded motion capture data of a walking person. The following invariants are taken into account.

- The ground plane is evenly textured and blends into a solid background, such that it gives the user some sense of perspective without distracting.
- Characters are fully textured and rendered using smooth shading. This provides the most realistic rendering of our character models, while maintaining the exact triangular shape used in the distance and collision computations.
- The characters are placed such that the (near) collision occurs in the vicinity of the origin.
- The camera is placed at an eye height of 1.75m and slightly looking downward as to show both characters from head to toe, mimicking the viewpoint of a human observer in a similar real-life situation. The downward angle is adjusted such that the point at 0.89m above the origin is at the center of the view.
- The camera's field of view is chosen to mimic a 50-mm lens on a 35-mm ("full-frame") camera, which is known to result in a perspective distortion similar to that of the human eye.
- Lights are attached to the camera at an offset; lighting is constant with respect to the camera angle.

Participants are presented with an online web-based questionnaire. Before starting the test, they are instructed that any physical contact between the displayed characters (including the slightest touch) is considered "a collision." Each participant is shown a scene, advancing to the next after the question "Do these characters collide?" is answered. The questions are binary; it is only possible to answer "yes" or "no." Answer buttons are always visible and can be used at any time. Participants have to click on their answer, and then on a confirmation button, which is placed equidistant to the "yes" and "no" buttons (see Figure 1). This ensures that the mouse has to travel a similar distance regardless of the answer to the previous question, preventing bias towards repeating answers.

The questionnaire, for both the pilot and main experiment, was open to any participant, who were sourced among colleagues, students, members of computer science and game development forums, and several other non-computer science forums. A small reward was raffled off among interested participants that completed the survey.

Four types of answers are considered: true positive (TP) when there was a collision and it was recognized as such; false positive (FP) when there was no collision but it was recognized as one; true negative (TN) when there was no collision and recognized as such; false negative (FN) when there was a collision but not recognized as one. Accuracy A is computed in a similar way as by Kulpa et al.,[9] as the fraction of correct answers

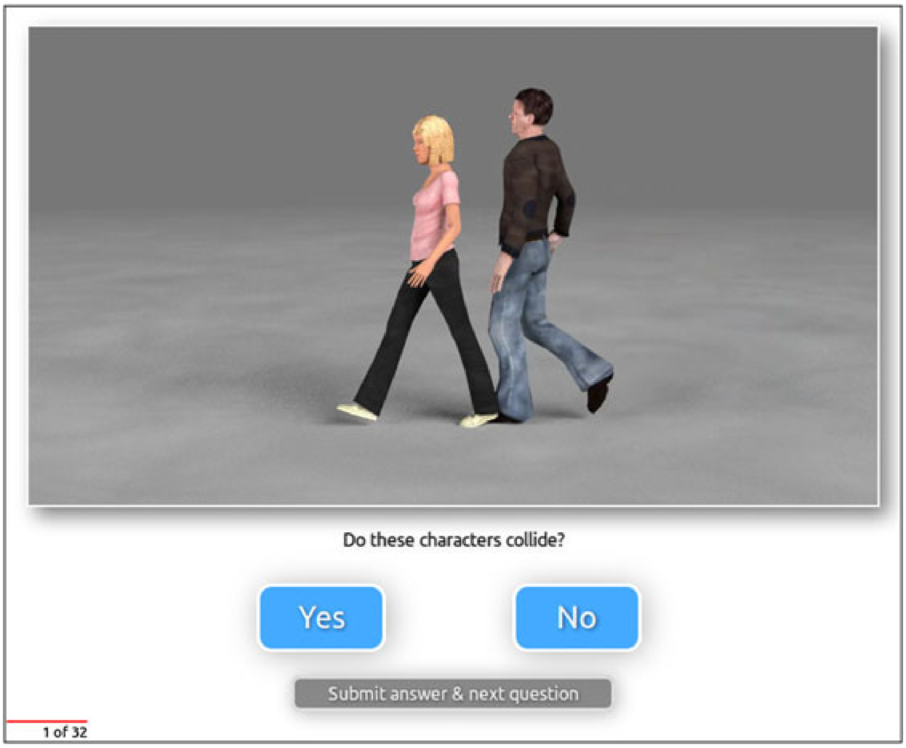

Do these characters collide?

FIGURE 1 Screenshot of the questionnaire

$$A = \frac{TP + TN}{TP + FP + TN + FN}$$

Absolute uncertainty (i.e., pure guesswork) would result in $A = 50\%$. With continuous data, absolute uncertainty would result in a normal distribution around $A = 50\%$; analysis of the standard deviation would be needed to determine whether the observed distribution differs significantly from pure guesswork. However, due to the binary nature of our data, the standard deviation contains no information regarding the spread of the answers.

4 | PILOT EXPERIMENT

In this section, we describe our pilot experiment. We investigate the ability of observers to recognize collisions between virtual characters in *static* situations. Using static images allows us to test a wider range of situations that are difficult to create in an animated context, especially given the requirements that there is only a single collision and that the forward velocity is more or less constant.

4.1 | Overview

This section describes the pilot experiment design, invariants, and variables. The invariants as described in Section 3 are taken into account, and shadow is not rendered at all.

The characters are placed on the ground (xy) plane such that the distance between their meshes is D_m meters. Both are initially placed in a random pose at the origin and then moved apart along the x-axis until the desired distance is obtained. A negative value for D_m models the penetration depth. For simplicity of computation, we use a reasonable approximation and define it as the minimum distance to travel along the x-axis in order to separate the two meshes.* To generate such cases, we use a two-step approach. First the characters are placed at $D_m = 0$ as described above, then they are both moved along the x-axis towards the origin by $D_m/2$; by moving both meshes, the collision will still be near the origin. The line segment L is defined by the closest points on the two meshes, hence $||L|| = \max(0, D_m)$ (assuming L is unique). When $D_m<0$, $||L||$ is a degenerate line segment, and defined as the point where the meshes touched in the first step of the two-step approach we described earlier.

The camera is placed at a random distance D_c to the centroid of L and a random angle, and an image is rendered. A selection of these images are then used in the user experiment; details are presented in Section 4.2. The *front* and *rear* character are defined respectively as the characters closest to and furthest from the camera, based on their root joint positions.

Our pilot experiment considers three variables. The first two variables are randomly sampled from a suitable distribution, and used as input to generate the images used in the experiment. The other variable is derived from the randomly generated scene.

- Mesh-mesh distance D_m was chosen uniformly from the interval $[-0.09, 0.15]$, in meters. We do not use any image with $|D_m|<0.001$.
- Camera distance $D_c \in [4,16]$, from camera to centroid of L, in meters. In the case of animated characters, Kulpa et al.[9] found that up to a certain projected size of the characters camera distance had little influence on accuracy. We are interested to see if this holds for static situations as well. This variable is chosen from an exponential distribution, such that more samples are chosen at smaller distances. When a character is close to the camera, perspective distortion is stronger and resulting effects are easier to measure. The lower bound is chosen such that characters fit entirely inside the camera frustum.
- Variable $\lambda \in [0, \infty)$ measures the length of the visible (i.e., not occluded by the front character) part of L, measured in meters. λ is undefined when the characters are colliding, as there is no visible gap between the characters in those cases. Note that this metric does not denote the gap between the *silhouettes* of the characters; there are many cases in which the visible part of L lies in front of the rear character, such as depicted in Figure 2a.

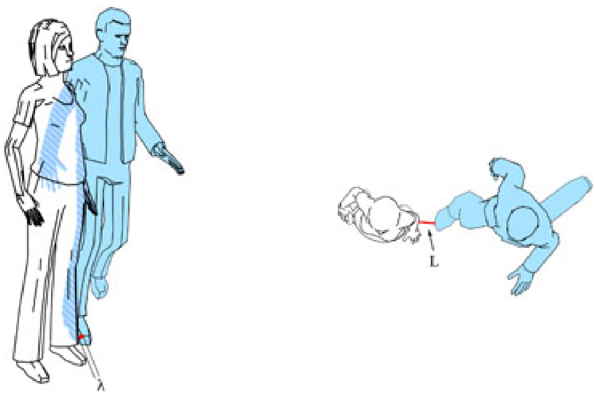
We want to investigate the role of these variables in the perception of collision detection. Given a configuration of two characters, variables D_m and λ are relatively hard to compute, because computing L is a non-trivial task. This means that these variables cannot be directly used in a crowd simulation system. Nevertheless, we suspect that they model important aspects of the perception of the observers. With respect to variable D_m, we expect the accuracy to be the lowest around $D_m = 0$, with a linear positive dependency between $|D_m|$ and the accuracy of observers. We expect a positive linear correlation between the accuracy and λ, as a more visible "gap" should result in higher accuracy.

Note that the camera angle is not directly part of the variables we consider. Even though the angle of the camera with respect to the walking direction of the characters has been shown in previous work to be relevant to perception.[9] our characters do not share a single direction, hence this metric loses its meaning.

4.2 | Experiment

To generate the images, the characters were posed using a randomly selected frame from a motion capture corpus

[*] Although theoretically there is the possibility that this metric is very different from the penetration depth, in our test cases this difference is only small.



(a) Front view as seen from the camera, showing λ as the visible part of line segment L in red. The striped area indicates the occluded part of the rear character.

(b) Top-down view, showing line segment L in red

FIGURE 2 Front and top-down view of the experiment setup

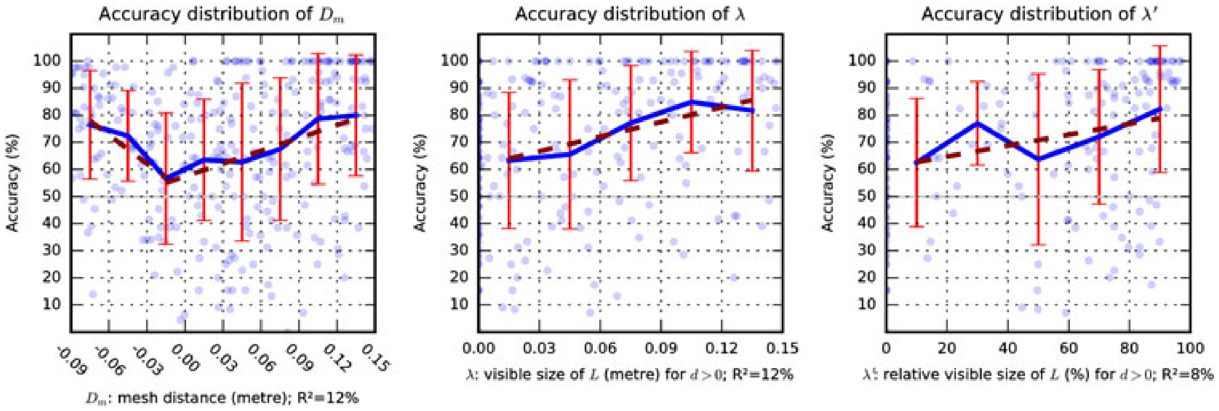


FIGURE 3 Plots of the results of the static experiment. Results are binned; the blue, continuous line shows the accuracy of each bin, with red error bars denoting the standard deviation. The red dotted line shows the linear fit of the accuracy. Note that the R^2 numbers relate only to a single parameter of our model with respect to the entire variance in the observations

consisting of stepping and walking motions. An additional random orientation around their up-axis prevented correlation between the test cases and the absolute orientation recorded in the motion capture lab.

The images are rendered at a resolution of 800 × 600 pixels.

Each variable's range is uniformly split up into eight bins, as shown in Figure 3. Each image is assigned a bin index for each of the three variables. The interval $[-0.09, 0.15]$ meters allows us to place $D_m = 0$ at a bin boundary, separating colliding and non-colliding images into different bins. A random sampling technique described by Wand and Straßer[23] is used to ensure at least 22 images per bin, resulting in a total of 373 images.

Participants are presented with an online web-based questionnaire, as described in Section 3. Each image is shown for 6s and is then hidden; this timeout ensures that all participants look at an image for a more or less equal duration. In order to prevent bias towards positive or negative answers, we include both colliding (i.e., D_m<0) and non-colliding (i.e., D_m>0) images in the experiment. An exact 50%/50% distribution for any single participant who completes the survey is ensured, and approximated for participants that do not.

Per participant, a random subset of the test cases is shown. This allows us to use a large test set without forcing participants to answer all 373 questions. Image selection is biased towards images in bins containing the least number of answers, providing a more even spread of answers over the bins than when uniformly selecting images.

4.3 | Results

A total of 212 actively participating users provided 9,179 answers, averaging 43 answers per participant. The accuracy over all participants was 72% for this experiment.

For each variable, a graph is shown in Figure 3. These graphs show the likelihood that the participants correctly identified the situation, averaged over the images in each bin. The solid blue graph shows the *average accuracy* per bin, with the error bars indicating the standard deviation. The scatter plot shows a dot for each image in the survey. The dark red dashed graph shows the *trend* and consists of one or two linear pieces.

To define the *trend* of the accuracy, we investigate, in order of simplicity, a linear function or a piecewise linear function. We accept the simplest function that describes the data well. An analysis of the variance shows how well the found trend fits the data (R^2).

As the projected size of λ is dependent on D_m, we also investigated the visible percentage $\lambda' = \lambda/D_m$ to remove this dependency. However, as can be seen in Figure 3, the results are less relevant than for λ ($R^2 = 8\%$).

The effect of camera distance D_c (not included in graph) is negligible, resulting in $R^2 = 0\%$ for camera distances up to 12m. This confirms that the findings by Kulpa et al. [9] are also applicable to static situations.

Computing the average over all participants, we see that 53% of the answers was "not colliding" and 47% was "colliding." The ratio of FP to FN provides information to the bias in the observations, and its significance can be computed using a binomial test. The *binomial test* is an exact test to compare the distribution of observations with an expected distribution, and can only be applied when there are two categories of observations (in this case, 'false positive' and 'false negative').[6] When there is no bias, the FP:FN ratio of N observations will show the same distribution as the ratio of heads:tails in N fair coin tosses; this is the null hypothesis. The binomial test results in the probability p that, given the observations, these distributions are indeed equal. When $p < 0.05$, it is 95% certain that the null hypothesis can be rejected, and we can interpret the FP:FN ratio as significantly different from fair coin tosses, and thus biased. We use a binomial test per participant, to compute the number of participants that are neutral, err towards false positives, or err towards false negatives. Using a 95% confidence interval, none of the participants had a significant bias towards false positives, that is, incorrectly answering "colliding." Fifty-eight percent of the participants did not show any bias, whereas 41% of the participants showed a significant bias towards false negatives, that is, incorrectly answering "not colliding."

By using the height of the (near) collision, we can separate the stimuli into "upper body" and "lower body." Figure 4 shows the distribution of the height of the collision, or the centroid of L in non-colliding cases, and confirms that such a distinction is sensible. We use k-means clustering ($k = 2$) to separate the test cases into "upper body" and "lower body" clusters, and apply the same analysis as before to each cluster individually.

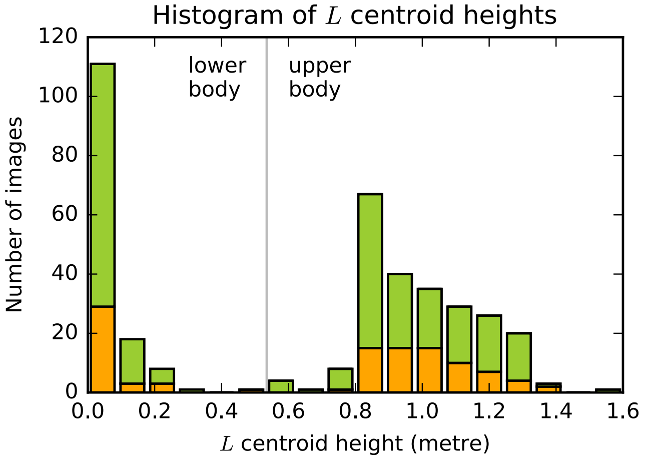


FIGURE 4 Histogram of the height of the static (near) collisions, clustered into "upper" and "lower" collisions

With $A = 71\%$ and $A = 73\%$ for respectively the upper and lower body, the overall accuracy is almost the same. However, the FN:FP ratio is different, with 63%:37% for the upper body and 47%:53% for the lower body. This difference is also reflected in the results of a binomial test. None of the participants showed a significant bias towards false positives, that is, incorrectly answering "colliding," both for the upper and lower body. Thirty-four percent and 17% of the participants, for respectively the upper and lower body, showed a significant bias towards false negatives, that is, incorrectly answering "not colliding." The remaining participants did not show significant bias.

These results may impact strategies for collision detection, as those are generally aimed at the prevention of false negatives; this analysis will be performed in the main experiment as well, to investigate whether the same bias towards false negatives is seen in animated situations.

5 | MAIN EXPERIMENT

In this section, we describe our main experiment, in which we investigate the ability of observers to recognize collisions between virtual characters in animated scenarios. Using animated characters, we aim for our results to be applicable to other animated situations, such as crowds of virtual characters.

5.1 | Overview and variables

This section provides an overview of the main experiment design. The two characters are animated using previously recorded motion capture data of a person walking in a straight line. Collision responses were not animated; participants could not discriminate colliding from non-colliding scenarios by looking at the animated behavior. Figure 5 shows an example still from one of the animations.

The same invariants as in Section 3 are taken into account, albeit with two differences with the pilot experiment. Firstly, to improve the perceived realism, and to visually ground the characters on the floor plane, shadows are rendered. These are very soft (i.e., no hard edges) by employing multiple, large light sources, preventing a second angle of view onto the scene. Secondly, since Kulpa et al.[9] showed that there is no statistical significance of the camera distance, the camera is placed at a fixed distance of 6.6m from the collision point.

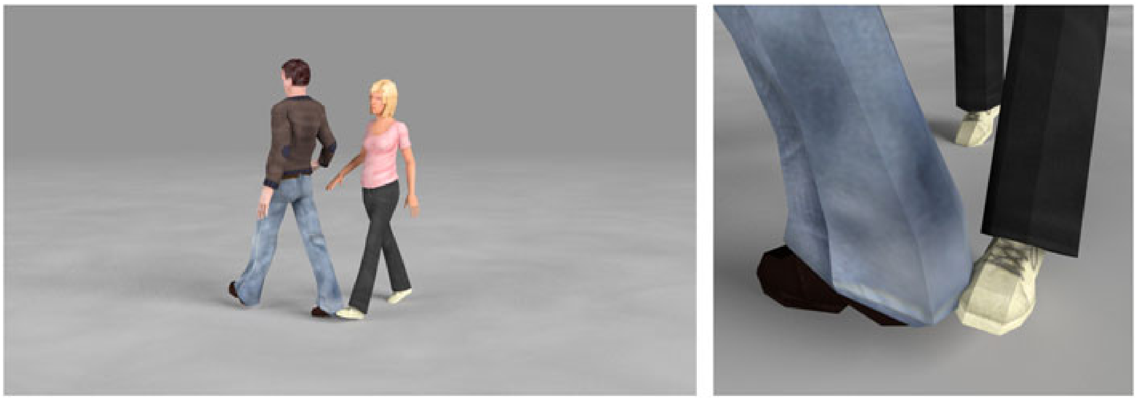


FIGURE 5 Still from one of the videos used in the animated experiment. Only 9% of the participants recognized the left scenario as a collision. The feet of the characters intersect, as can be seen from the side view on the right

The characters are placed on the ground (XY) plane, such that their paths cross at an angle α. By modifying this angle, the starting positions, and animation offsets, a total of 16 colliding and 16 non-colliding scenarios were constructed. To allow reasoning about *the* collision, we make sure that for the colliding cases there was only a single, continuous time interval in which the characters were intersecting. As a result, $\alpha = 0$ could not be investigated, as it would be impossible to create a scenario with a single collision of the intended severity.

To give the best view of the (near) collision, the camera is placed on either the positive or negative side of one of the two bisectors of the character's paths. For each video, we manually select which of those four possible positions provides the best view. This provides us with a worst case scenario, as when looking at animated characters in general, often the user will not have an unobstructed view of the collision.

Our experiment considers four variables, defined below. The first two are selected from a set of predefined values. The character animation is adjusted as described earlier, to produce a video that adheres to those parameters. The second two parameters are derived from this animation, and allow us to perform more statistical analyses on our data in order to find out which component is important for the recognition of collisions. The variables are defined as follows.

- Character angle $\alpha \in \{45{,}90{,}135{,}180\}$ degrees. This defines the angle between the forward vectors of the characters.
- The severity S of the (near) collision labeled as LOW, MODEST, CONSIDERABLE or HIGH, and expressed either as intersection volume integrated over time (I_V) when colliding, or as the minimum mesh distance (D_m) otherwise. See Table 1 for the values used; each scenario used one of the displayed values, precise up to one decimal.
- Collision duration τ is derived from the animation created to obtain the first two parameters. This variable is only defined for colliding videos.
- Average intersection volume $I_A = I_V/\tau$. This variable is defined only for colliding videos.

TABLE 1 Severity labels for the colliding and non-colliding cases, based on a small pilot experiment

Label	Colliding: I_V	Label	Non-colliding: D_m
LOW	0.5 cm^{3}s	LOW	0.5 cm
MODEST	12.5 cm^{3}s	MODEST	1.0 cm
CONSIDERABLE	67.2 cm^{3}s	CONSIDERABLE	3.0 cm
HIGH	132.0 cm^{3}s	HIGH	5.0 cm

The severity labels have different meaning for colliding and non-colliding cases. In the colliding cases, there is a temporary overlap between the two characters. This is expressed in the size of the intersecting volume (in cm^3) integrated over time (in seconds), giving us the integral I_V in cm^{3}s. Both aspects (size and duration) are important to quantify the potential recognizability of the collision, as even a small intersection will be seen when existing for a long enough time. In the non-colliding cases, the severity was defined as the minimum distance between the meshes D_m, which is identical to the D_m parameter as defined in Section 4.1, except that now this minimum is taken over the entire spatio-temporal domain.

To find a suitable range for the collision severity, we have conducted a small pilot experiment with three participants. It took the same form as the actual experiment, and used the following values for the variables:

- $\alpha \in \{45{,}90{,}135{,}180\}$ degrees
- $I_V \in \{0.5{,}30{,}150{,}300\}$cm^{3}s
- $D_m \in \{0.00{,}\pm0.05{,}\pm0.10{,}\pm0.20\}$ meters

This small pilot showed that for the larger values of the collision severities I_V and D_m, in respectively the colliding and non-colliding cases, it was very easy to recognize a (non-)collision. Removing these values allows us to have a finer granularity in the lower, more interesting range, without increasing the number of required videos. As stated before, the values used for the final experiment are shown in Table 1.

5.2 | Image generation and questionnaire

This section describes the details of the main user experiment. Two textured character models were selected for the experiment; every image uses the same two models with the same textures to prevent dependence on the appearance. We ensured a high contrast between arms and legs of both characters, to make distinguishing the two characters as easy as possible.

We used four slightly different walking motions to reduce the learning effect and ensured that the two characters never used the same motion in the same video. For each combination of variables, and for both colliding and non-colliding, we generated a video at a standard resolution of 1280 × 720 pixels at 30 frames per second. Each video was 2.5s long. The time of the (near) collision was randomly chosen between the 50% and 75% mark, to prevent a learning effect. The starting position of the character and the offset into the walking animation were chosen manually, in order to be able to ensure a (near) collision of the intended severity.

Participants are presented with an online web-based questionnaire (see Section 3 and Figure 1). Each participant is shown all 32 video clips in random order. Each clip is played once, advancing to the next video after the question “Do these characters collide?” is answered.

5.3 | Results

The data of the animated experiment is based on the answers of 164 participants, each providing exactly 32 answers. In total, 195 people participated; 30 did not complete the survey, and one participant completed the survey on a mobile device. That mobile user’s data was not considered, as the small size of the screen makes recognizing collisions harder. The accuracy over all participants was 68% for this animated experiment. Note that the difference between this result and the 72% accuracy observed in the pilot experiment does *not* imply that people are better at recognizing static scenarios, because there are more differences between the experiments than just being static or animated. The hardest to recognize collision is shown in Figure 5.

For the colliding cases, that is, the scenarios in which $FP = TN = 0$, the accuracy is 58%. For the non-colliding cases, that is, in which $TP = FN = 0$, the accuracy is 79%.

In order to understand the relation of our variables α and S, and the expected accuracy $E[A]$, we apply linear regression analysis.[10]

The analysis is performed on all answers, and not just on the averages per bin; this implicitly takes variance of the answers into account. Firstly, we linearize our input by finding as simple as possible functions $f_1(\alpha)$ and $f_2(S)$. Secondly, we use a statistical software package to find the best-fitting B_0, B_1, B_2,

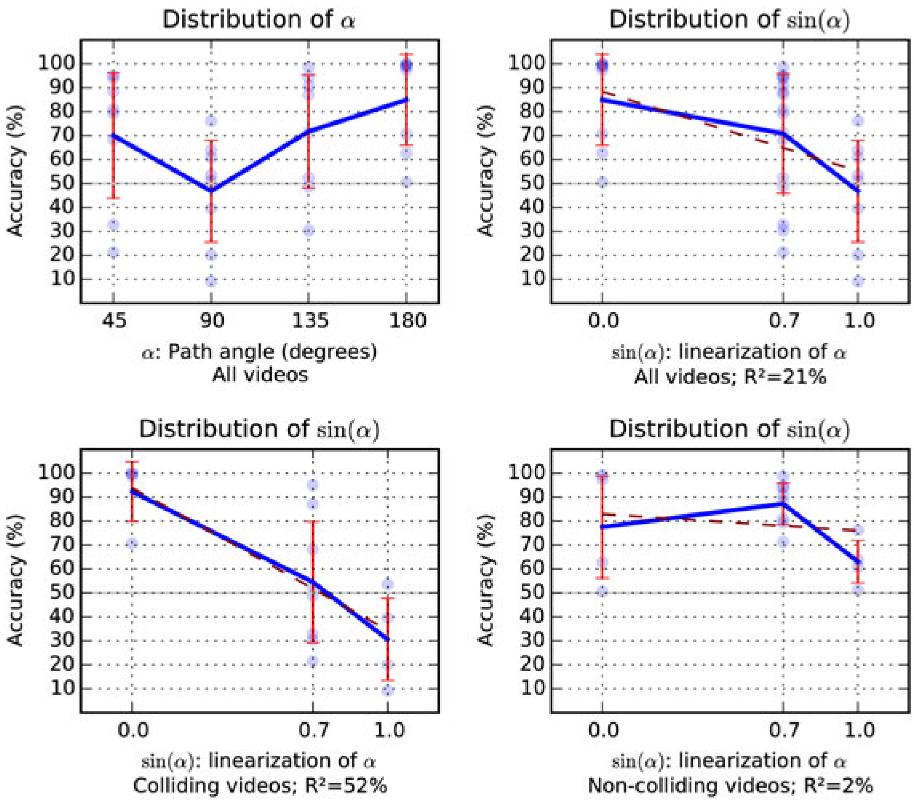


FIGURE 6 Plots of the contribution of α to the results of the animated experiment. The blue line shows the accuracy of each bin, with red error bars denoting the standard deviation. The dashed red line indicates the linear fit. Note that the R^2 numbers relate only to a single parameter of our model with respect to the entire variance in the observations

TABLE 2 Linear regression model for the colliding cases; R^2 = 78%

	Coefficients		Standardized	Significance	
	B	Std. Err.	coefficients	t	p
(Constant)	$B_0 = 0.89$	0.10		8.98	0.000
$\sin(\alpha)$	$B_1 = -0.73$	0.13	$\beta_1 = -0.77$	−5.54	0.000
I_V	$B_2 = 0.01$	0.00	$\beta_2 = 0.02$	3.21	0.008
$\sin(\alpha) \times I_V$	$B_3 = 0.00$	0.00	$\beta_3 = 0.52$	−1.95	0.074

TABLE 3 Linear regression model for the colliding cases; R^2= 80%. Interaction between the variables is insignificant ($p > 0.7$) and not included in this table

	Coefficients		Standardized	Significance	
	B	Std. Err.	coefficients	t	p
(Constant)	$B_0 = 0.72$	0.10		7.58	0.000
$\sin(\alpha)$	$B_1 = -0.63$	0.11	$\beta_1 = -0.77$	5.88	0.000
I_A	$B_2 = 0.00$	0.00	$\beta_2 = 0.02$	0.16	0.879
τ	$B_3 = 1.47$	0.45	$\beta_3 = 0.52$	3.29	0.007

and B_3 such that

$$E[A] = B_0 + B_1 f_1(\alpha) + B_2 f_2(S) + B_3 f_1(\alpha) f_2(S) \qquad (1)$$

Because S is expressed differently for colliding and non-colliding scenarios, we perform the linear regression method for each separately.

Before applying the linear regression analysis, we need to find suitable functions $f_1(\alpha)$ and $f_2(S)$. The α graph in Figure 6 shows a more or less sine-like shape, which could indicate a relation between $E[A]$ and the size of the projection of one of the trajectories onto the other. We choose $f_1(\alpha) = \sin(\alpha)$; as rotations are periodic, we expect the influence of α on $E[A]$ to be periodic as well, supporting the choice for a periodic linearization.[†]

The I_V graph is fairly linear, except for the data point at LOW. We use $f_2(I_V) = I_V$, but we may consider a different linearization in the future; we will get back to this in Section 6. We feel that the accuracy distribution curve of D_m is sufficiently linear, resulting in $f_2(D_m) = D_m$.

Results of the linear regression are shown in Tables 2–5, with the B_i coefficients in the second column. The R^2 value mentioned in each caption denotes the percentage of the variance explained by these models. Any row with $p < 0.05$ is considered *significant*, and with $p < 0.01$ considered *strongly significant*.

When the characters are colliding, a linear combination of $\sin(\alpha)$ and I_V predicts 78% of the variance in A (see Table 2). The interaction between the two variables is not significant ($p = 0.07$). Because I_V is the volume of the intersection integrated over time, we can split its value into average volume I_A and duration τ, to investigate which aspect is more important to the correct classification of the video by observers. This results in the model shown in Table 3, with R^2 = 80%.

TABLE 4 Linear regression model based on $\sin(\alpha)$ and $D_m$for the non-colliding cases; R^2 = 47%

	Coefficients		Standardized	Significance	
	B	Std. Err.	coefficients	t	p
(Constant)	$B_0 = 0.59$	0.11		5.32	0.000
$\sin(\alpha)$	$B_1 = 0.12$	0.16	$\beta_1 = 0.28$	0.79	0.446
D_m	$B_2 = 0.10$	0.04	$\beta_2 = 1.12$	2.78	0.017
$\sin(\alpha) \times D_m$	$B_3 = -0.08$	0.05	$\beta_3 = -0.76$	−1.54	0.148

TABLE 5 Linear regression model based on $D_m$for the non-colliding cases; R^2= 34%

	Coefficients		Standardized	Significance	
	B	Std. Err.	coefficients	t	p
(Constant)	$B_0 = 0.66$	0.06		11.08	0.000
D_m	$B_1 = 0.05$	0.02	$\beta_1 = 0.59$	2.70	0.017

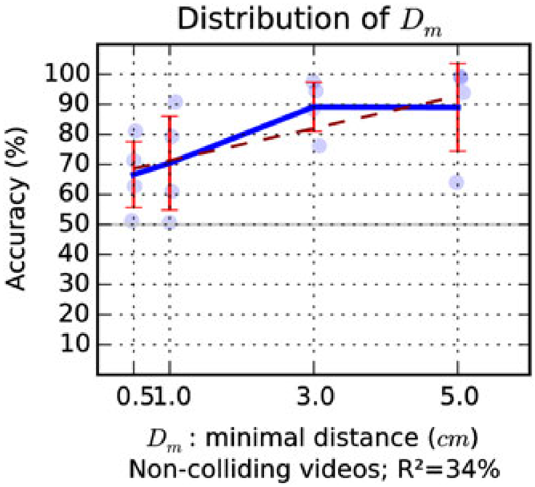


FIGURE 7 Plot of the results of the animated experiment. The blue line shows the accuracy of each bin, with red error bars denoting the standard deviation. The dashed red line indicates the linear fit. Note that the R^2 number relate only to a single parameter of our model with respect to the entire variance in the observations

Even though there is variation in A that we did not capture in our model, such as the characters' exact poses at the moment of collision, our model is significant to A. The interaction between I_A and τ is expressed as I_V, and is not included in this analysis due to its insignificance.

Interestingly, with $\beta_1 = -0.77$, the angle between the characters is the most important factor. The sign of β_1 indicates a negative correlation, as the minimum accuracy was measured at $\alpha = 90^o$. The duration of the collision is slightly less important, with $\beta_3 = 0.52$. The average volume of the collision is insignificant ($p = 0.879$).

In the non-colliding cases, the model based on $\sin(\alpha)$, D_m, and their interaction seems to predict 47% of the variance in A (see Table 4). However, because a linear model always produces a better fit when there are more parameters, we have to remove the non-significant parameters from the analysis.[‡]

[†] We also investigated $f_1(\alpha) = \alpha$, $f_1(\alpha) = \alpha^2$, $f_1(\alpha) = \cos(\alpha)$, and $f_1(\alpha) = \sin(\alpha + \pi/4)$, but all these variants resulted in a lower R^2 than $f_1(\alpha) = \sin(\alpha)$.

[‡] This was not necessary for the analysis shown in Table 3; due to the very small β_2, the outcome would not change significantly.

This is also reflected in Figure 6, which shows how $\sin(\alpha)$ alone explains 52% of the variance of the colliding cases, but only 2% of the non-colliding cases. The model based on only D_m, shown in Table 5 and Figure 7 results in $R^2 = 34\%$. Even though this R^2 is moderate, the results are significant.

Computing the average over all participants, we see that 60% of the answers was "not colliding" and 40% "colliding." To obtain more detail of the nature of this apparent bias, we investigate the ratio of false positives and false negatives for each participant. A two-tailed binomial test using a 95% confidence interval (as described in Section 4.3) showed that 72 of the participants did not have a bias, and no participants had a bias towards false positives. The remaining 92 participants had a bias towards false negatives; in other words, the majority of the participants significantly erred towards answering "not colliding."

6 | DISCUSSION

Looking at our findings, we can conclude that in general the subjects were better at recognizing non-collisions than collisions.

The results from both experiments show the same trends, and although they cannot directly be mapped onto each other, this could indicate that trends observed in static situations may be applicable to animated scenarios as well.

In our main experiment, an intersection volume integral of $0.5\text{cm}^3\text{s}$ resulted in an accuracy of only 33%. Apparently, for the average observer, it is the most difficult to classify a scenario as "colliding" or "not colliding" when there is a small amount of interpenetration. This is also observed in the pilot experiment, where we see a remarkable dip in accuracy in the interval $D_m \in [-0.03, 0.00)$. The bias towards answering "not colliding," observed in both experiments, corroborates these observations. This knowledge may be used to speed up collision detection algorithms. A simplified version of the mesh could be created, taking care that it is an "inner approximation" bounded by the original mesh. By ensuring a Hausdorff distance[22] of at most 1.5 cm the total penetration of two such meshes would be at most 3.0 cm and fall within the interval of minimal accuracy. The algorithm to create a simple mesh that meets those requirements, and the effect on both the perception of collisions and the performance of collision detection, is an interesting open problem. These observations also seem to indicate that, for collision detection between humanoid shapes, a bounding volume collision detection scheme may not be the best choice. Employing a *bounded* volume method representing an inner approximation could be more efficient, and a better match for our perception.

Intersections are allowed in certain commercial crowd simulation systems, such as the implementation in IO

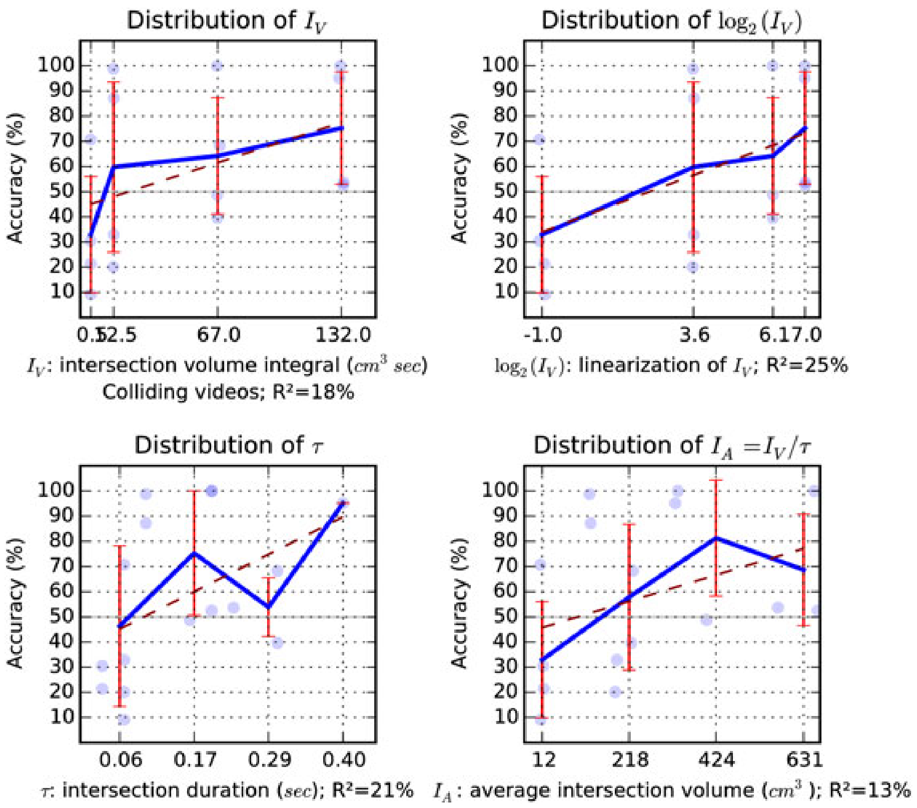


FIGURE 8 Plots of the contribution of I_V, I_A, and τ to the results of the animated experiment. The blue line shows the accuracy of each bin, with red error bars denoting the standard deviation. The dashed red line indicates the linear fit. The graph of $\log_2(I_V)$ is discussed in Section 6. Note that the R^2 numbers relate only to a single parameter of our model with respect to the entire variance in the observations

Interactive's *Hitman: Absolution*; apparently large game companies assume that people do not mind such partial intersections.[23] Such an approach also allows for denser crowds, simply by decreasing the personal radius of the characters, without sacrificing too much believability. The low importance of the intersection volume I_V (see Table 2), coupled with the high importance of the angle α, also suggests varying the effective personal radius for collision checking based on the angle between the paths of the checked characters.

From the λ accuracy graph, it is clear that the more visible the gap between the characters, the easier it is to see that they do not collide. Note that this metric does not denote the gap between the *silhouettes* of the characters – there are many cases in which the visible part of L lies in front of the rear character.

Alternate linearizations for α, I_V, and D_m might produce a better fitting model. The I_V accuracy graph resembles a logarithmic curve, so we have also investigated $f_2(I_V) = \log_2(I_V)$, resulting in the curve shown in the top left of Figure 8. This model is a tighter fit for the data ($R^2 = 82\%$ for the entire model, instead of 80%). However, even though visually a logarithm may fit the graph well, this does not imply that it is the best model to use. For this reason, we have kept the linearization simple, and leave more complex linearizations to future research.

We would have liked to apply the upper body and lower body analysis we performed in the pilot experiment to the results of the main experiment. However, for every pair of parameters (α, S), there was only one sample, that is, only an upper or a lower body collision. Furthermore, there is an imbalance between the number of colliding and non-colliding cases for each body half. It would be interesting to study these differences between perception of upper and lower body collisions, along with possible influential factors, such as timing and volume of the collision. We leave such a study to future work.

7 | CONCLUSION

In this paper, we have conducted a perceptual experiment to determine the accuracy of human observers in determining whether two virtual characters collide. We have identified an asymmetry in the recognition of collisions, a critical penetration depth interval where the accuracy is minimal, and proposed a level of detail technique that utilizes this knowledge to speed up collision detection. New collision response criteria that increase performance and allow denser crowds by focusing on pairs of characters have been introduced.

Care should be taken in those cases where crowd behavior is changed based on any camera-related metric. When crowd simulation is used to mimic real humans, for example, to evaluate evacuation scenarios, such view-dependent behavior will change the outcome of the simulation. When crowd simulation is used in games, view-dependent behavior could be exploited to gain unfair advantage over other players. For example, one could turn off collision detection of crowd agents when they are not in view of the player; this would make traversing a crowd easier when walking backward than when walking forward.

Simplified shapes are often used in physics simulation software. Future research could investigate whether the results in this paper are applicable only to humanoid shapes or generalize to other objects or even abstract geometric shapes.

In our surveys, we have not rendered crisp shadows and ambient occlusion. This simplifies rendering; shadowless rendering is also used in commercial applications[23] to enable real-time rendering of crowds. It would be interesting to see the effect of different types of shading and lighting on the perception of the crowd in general and collisions in particular.

The backgrounds were rendered as simple as possible, to ensure our results depend only on the two virtual characters and the camera position. The effects of the background behind the characters, especially when visible in the space between the characters, is still an open research question.

We observed an asymmetry in the recognition of collisions, and a bias towards answering "not colliding." These effects could have several causes.

Firstly, the characters did not employ a collision response animation. Because of this, and because real humans do not intersect each other when colliding, the non-colliding and colliding scenarios could be classified as respectively realistic and unrealistic, causing this bias towards realistic scenarios.

Secondly, participants may have focused on the spot where they anticipated a collision. In cases where they anticipated incorrectly, such focus may have caused them to miss the collision. Because the other way around cannot occur, this likely contributed to the observed bias.

Thirdly, we also observed that most of the collisions occurred between the hands or the feet. This was likely caused by the use of a simple walk animation that was not adapting to the proximity of the other character. Real humans would probably be able to slightly change their hand or foot position to avoid a collision without changing their own global position or trajectory. Expecting such behavior may have also accounted for the bias towards answering "not colliding." This bias could be used by choosing a representation that allows for some undetected collisions.

In future work, it would also be interesting to see how collision avoidance and response animations influence this bias specifically, and the perception of collisions in general.

REFERENCES

1. O'Sullivan C, Dingliana J. Collisions and perception. ACM Trans Graph. 2001;20(3):151–168.

2. O'Sullivan C, Radach R, Collins S. A model of collision perception for real-time animation. In: Magnenat-Thalmann N., Thalmann D., editors. Computer Animation and Simulation '99, Eurographics. Springer: Vienna; 1999. p. 67–76.

3. Ennis C, Hoyet L, Egges A, McDonnell R. Emotion capture: emotionally expressive characters for games. Proceedings of Motion on Games, MIG '13. ACM: New York, NY, USA; 2013. p. 31:53–31:60.
4. Ennis C, McDonnell R, O'Sullivan C. Seeing is believing: body motion dominates in multisensory conversations. ACM SIGGRAPH 2010 Papers, SIGGRAPH '10. ACM New York, NY, USA; 2010. p. 91:1–91:9.
5. McDonnell R, Ennis C, Dobbyn S, O'Sullivan C. Talking bodies: sensitivity to desynchronization of conversations. ACM Trans Appl Percept. 2009;6(4):22:1–22:8.
6. Cohen J, Varshney A, Manocha D, Turk G, Weber H, Agarwal P, Brooks F, Wright W. Simplification envelopes. Proceedings of the 23rd annual conference on computer graphics and interactive techniques, SIGGRAPH '96. ACM: New York, NY, USA; 1996. p. 119–128.
7. DeLucia PR. Effects of size on collision perception and implications for perceptual theory and transportation safety. Curr Dir Psychol Sci. 2013;22(3):199–204.
8. Dingliana J, O'Sullivan C. Graceful degradation of collision handling in physically based animation. Comput Graph Forum. 2000;19(3):239–248.
9. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions: John Wiley & Sons: New York; 2013.
10. Gauss CF. Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Reissue edition (May 19, 2011), Cambridge Library Collection - Mathematics: Cambridge, UK: Cambridge University Press; 1809.
11. Hausdorff F. Grundzüge Der Mengenlehre. Veit and Company: Leipzig; 1914.
12. Hubbard PM. Collision Detection for Interactive Graphics Applications. *Ph.D. Thesis*, Providence, RI, USA: Brown University; 1994.
13. Kulpa R, Olivierxs AH, Ondřej J, Pettré J. Imperceptible relaxation of collision avoidance constraints in virtual crowds. ACM Trans Graph. 2011;30(6):138:1–138:10.
14. LaValle SM. Planning Algorithms: Cambridge University Press: Cambridge; 2006.
15. Luebke D, Watson B, Cohen JD, Reddy M, Varshney A. Level of Detail for 3d Graphics. Elsevier Science Inc.: New York, NY, USA; 2002.
16. O'Sullivan C, Cassell J, Vilhjálmsson H, Dingliana J, Dobbyn S, McNamee B, Peters C, Giang T. Levels of detail for crowds and groups. Comput Graph Forum. 2002;21(4):733–741.
17. Otaduy MA, Lin MC. CLODs: Dual hierarchies for multiresolution collision detection. Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, SGP '03. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2003. p. 94–101.
18. Stüvel SA, Ennis C, Egges A. Mass population: Plausible and practical crowd simulation, Chapter 6. In: Angelides MC, Agius H, editors. Handbook of Digital Games. Piscataway, NJ, USA: Wiley-IEEE Press; 2014. p. 146–174.
19. Yoon SE, Salomon B, Lin M, Manocha D. Fast collision detection between massive models using dynamic simplification. Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, SGP '04. ACM: New York, NY, USA; 2004. p. 136–146.
20. Hoyet L, McDonnell R, O'Sullivan C. Push it real: perceiving causality in virtual interactions. ACM Trans Graph. 2012;31(4):90:1–90:9.
21. Olivier AH, Ondřej J, Pettré J, Kulpa R, Crétual A. Interaction between real and virtual humans during walking: perceptual evaluation of a simple device. Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization, APGV '10. ACM: New York, NY, USA; 2010. p. 117–124.
22. Stüvel SA, Magnenat-Thalmann N, Thalmann D, Egges A, van der Stappen F. Hierarchical structures for collision checking between virtual characters. Comput Animat Virtual Worlds. 2014;25(3-4):333–342.
23. Wand M, Straßer W. Multi-resolution rendering of complex animated scenes. Comput Graph Forum. 2002;21(3):483–491.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Stüvel, S. A., van der Stappen, A. Frank, and Egges, A. (2016), Perception of collisions between virtual characters, *Comp. Anim. Virtual Worlds*. doi: 10.1002/cav.1728

AUTHORS BIOGRAPHIES

Sybren A. Stüvel obtained his PhD degree at the Virtual Human Technology Lab in the Department of Information and Computing Sciences, Utrecht University, the Netherlands. His main research topic is the animation of dense crowds of virtual characters. He obtained his Master′s degree in Computer Science in 2010 at the same research group, and developed a method for the generation of human locomotion based on footstep positions. He is co-author of the IEEE Handbook of Digital Games and is part of the COMMIT/ project.

A. Frank van der Stappen received the M.Sc. degree from Eindhoven University of Technology, Eindhoven, the Netherlands, in 1988 and the PhD degree from Utrecht University, Utrecht, the Netherlands, in 1994. He is currently an Associate Professor with the Department of Information and Computing Sciences, Utrecht University. His research interests include manipulation, motion planning, grasping, simulation, animation, and geometric algorithms.

Arjan Egges is an Associate Professor at the Virtual Human Technology Lab in the Department of Information and Computing Sciences, Utrecht University in the Netherlands. He obtained his PhD at MIRALab - University of Geneva, Switzerland on the topic of emotion and personality models, in combination with automatically generated face and body motions using motion capture data. His current research focuses on crowd animation and motion perception as a part of the COMMIT and TARDIS projects. He teaches several courses related to games and computer animation and is one of the founders of the annual Motion in Games conference.