



# Perception of Motion Variations in Large-Scale Virtual Human Crowds

Robin Adili, Benjamin Niay, Katja Zibrek, Anne-Hélène Olivier, Julien Pettré,  
Ludovic Hoyet

## ► To cite this version:

Robin Adili, Benjamin Niay, Katja Zibrek, Anne-Hélène Olivier, Julien Pettré, et al.. Perception of Motion Variations in Large-Scale Virtual Human Crowds. MIG 2021 - 14th Annual ACM SIGGRAPH Conference on Motion, Interaction and Games, Nov 2021, Virtual Event Switzerland, France. pp.1-7, 10.1145/3487983.3488288 . hal-03468589

**HAL Id: hal-03468589**

**<https://inria.hal.science/hal-03468589>**

Submitted on 7 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Perception of Motion Variations in Large-Scale Virtual Human Crowds

Robin Adili  
robin.adili@inria.fr  
Inria, Univ Rennes, CNRS, IRISA  
Rennes, France

Benjamin Niay  
benjamin.niay@inria.fr  
Inria, Univ Rennes, CNRS, IRISA  
Rennes, France

Katja Zibrek  
katja.zibrek@inria.fr  
Inria, Univ Rennes, CNRS, IRISA  
Rennes, France

Anne-Hélène Olivier  
anne-helene.olivier@inria.fr  
Inria, Univ Rennes, M2S, CNRS, IRISA  
Rennes, France

Julien Pettré  
julien.petre@inria.fr  
Inria, Univ Rennes, CNRS, IRISA  
Rennes, France

Ludovic Hoyet  
ludovic.hoyet@inria.fr  
Inria, Univ Rennes, CNRS, IRISA  
Rennes, France



**Figure 1: Examples of virtual crowds stimuli: (left) 250 characters animated with 1 female and 1 male motion, (centre) 500 characters animated with 25 female and 25 male synthetic motions, (right) 1000 characters animated with 500 female and 500 male synthetic motions (i.e. a unique motion per character).**

## ABSTRACT

Virtual human crowds are regularly featured in movies and video games. With a large number of virtual characters each behaving in their own way, spectacular scenes can be produced. The more diverse the characters and their behaviors are, the more realistic the virtual crowd is expected to be perceived. Hence, creating virtual crowds is a trade-off between the cost associated with acquiring more diverse assets, namely more virtual characters with their animations, and achieving better realism. In this paper, our focus is on the perceived variety in virtual crowd character motions. We present an experiment exploring whether observers are able to identify virtual crowds including motion clones in the case of large-scale crowds (from 250 to 1000 characters). As it is not possible to acquire individual motions for such numbers of characters, we rely on a state-of-the-art motion variation approach to synthesize unique variations of existing examples for each character in the crowd. Participants then compared pairs of videos, where each character was animated either with a unique motion or using a subset of these motions. Our results show that virtual crowds with

more than two motions (one per gender) were perceptually equivalent, regardless of their size. We believe these findings can help create efficient crowd applications, and are an additional step into a broader understanding of the perception of motion variety.

## CCS CONCEPTS

- **Computing methodologies** → **Motion capture; Perception;**
- **Human-centered computing** → *User studies.*

## KEYWORDS

Virtual Crowds, Motion Synthesis, Perception, Character Animation

### ACM Reference Format:

Robin Adili, Benjamin Niay, Katja Zibrek, Anne-Hélène Olivier, Julien Pettré, and Ludovic Hoyet. 2021. Perception of Motion Variations in Large-Scale Virtual Human Crowds. In *Motion, Interaction and Games (MIG '21)*, November 10–12, 2021, Virtual Event, Switzerland. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3487983.3488288>

## 1 INTRODUCTION

Crowds are a common occurrence in human societies. When they appear in visual entertainment, the sheer number of characters on screen, each ideally looking and behaving differently while interacting with each other, produces spectacular and lively scenes. Until recent decades, hundreds if not thousands of extras needed to be hired and coordinated to capture a scene featuring a large crowd. With the development of computer graphics and visual

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

MIG '21, November 10–12, 2021, Virtual Event, Switzerland

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-9131-3/21/11...\$15.00

<https://doi.org/10.1145/3487983.3488288>

effects, crowds can nowadays be recreated in a virtual environment, and composited into movies or included in interactive applications.

However, what makes crowds appealing also makes them difficult to recreate in virtual environments. Animating and rendering such a large number of characters is often resource intensive, as every character of the simulated crowd is ideally unique in appearance and behaviour. Dedicated techniques are therefore required, e.g., for simulating crowd trajectories [van Toll and Pettré 2021], creating distinct character appearances [McDonnell et al. 2008, 2009b], animating the characters [Hoyet et al. 2016], rendering the crowd [Beacco et al. 2016; Dobbyn et al. 2005], etc. However, as capturing or creating the appearance and animation of every single character would defeat the purpose of crowd simulation, virtual crowds are usually created from a smaller set of characters, by using the same characters multiple times, potentially using procedural variations [McDonnell et al. 2008, 2009b]. In this context, identical instances of characters are referred to as clones. When their motion is identical, but their appearance is different, they are referred to as motion clones. The smaller the set of original characters and individual motions, the more clones will be displayed and the more uniform the crowd will look, which may impair visual realism.

In this paper, we focus on the perception of motion variety in virtual crowds. Introducing more varied motions comes at a cost, both in terms of acquisition and runtime performance. Hence, knowing just how many motions is enough for a crowd to be realistic is a valuable information when trying to design efficient crowd applications. In the context of small-scale crowds (maximum 24 characters), Pražák and O’Sullivan [2011] showed that using only three different motions can be sufficient to make a crowd look varied. However, it is unclear whether these results would remain valid for large scale crowds, as displaying more characters could make clones easier to spot or influence their perception in other, unforeseen manners. Our goal is therefore to investigate the perception of motion clones in large-scale crowds (up to 1000 characters). Further motivation is provided by the findings of McDonnell et al. [2008], which showed that motions clones that appeared more often were also more quickly spotted. However, as it is still impossible to acquire unique individual motions for such numbers of characters, we are also interested in exploring whether statistics-based methods for creating motion variations can be used to increase the visual variety and realism of such large-scale crowds.

To explore these questions, we generated virtual character crowd scenes using full-body walking motions synthesised using a statistics-based approach [Ma et al. 2010] from a motion captured database of 43 actors (20F, 23M). We then conducted an online experiment where participants were presented with video pairs of crowd scenes. On one video of each pair, every character was animated with a unique synthesized motion, while on the other video the characters’ animations were a subset of these motions. The participants’ task was to select the video which they considered had the most unique animations. We also explored whether the number of characters present in the crowd (250, 500 or 1000) affected the perception of motion variations. In our results, crowds with more than 2 motions (one per character gender) were not reliably distinguished from fully varied crowds of any size. As most of the stimuli were found to be perceptually equivalent, our results suggest that even in large-scale crowd little motion variety is required. Some other,

unaccounted factors like viewer’s point of view, motion distinctiveness, or character trajectories might play a hidden role, and we believe that future works could follow ours in these directions.

## 2 RELATED WORK

*Perception of human motion.* has been a very active field of research since the pioneer work of Johansson [1973]. Using a stimulus known as the “point-light walker”, he demonstrated that a few bright spots positioned on a person’s body against a contrasted background are enough to be perceived as a human motion. Following work showed that it is also possible to recognize from such stimuli: gender [Kozlowski and Cutting 1977; Troje 2002], motions of a friend [Cutting and Kozlowski 1977], or one’s own walking pattern [Beardsworth and Buckner 1981]. Since, the perception of human motion has also largely interested the Graphics community, especially as virtual humans have been demonstrated to convey more information than point light display [Chaminade et al. 2007]. E.g., studies involving virtual characters have explored the perception of motion artifacts [Hodgins et al. 2010; Pražák et al. 2011], of emotions conveyed by virtual characters [McDonnell et al. 2009a; Zibrek et al. 2015], of physical interactions between characters [Hoyet et al. 2012; Vicovaro et al. 2014], or of the ability to differentiate motions of different actors [Hoyet et al. 2013].

*Perception of variety in crowds.* is more directly related to the topic of this paper, and O’Sullivan’s survey [2009] provides a good overview of the problems involved. Introduction of variety in character features of a crowd scene is motivated by the viewers’ attention to those features, and their tendency to remember and recognise them at a later time or on other characters. Various works have explored this topic from a perceptual standpoint, with a focus on one or multiple character features, by conducting perceptual experiments where more or less varied virtual crowd scenes are shown to participants who evaluate their realism.

McDonnell et al. [2008] were the first to explore the effect of crowd variety both in terms of appearance and motion. In the stimuli they showed participants, only one motion or appearance was cloned in a crowd of 20 characters. Their results showed that motion clones are significantly more difficult to spot than appearance clones, which can be masked by color variation, random orientation, and motion. In following work, McDonnell et al. [2009b] further explored which body parts of virtual characters are most looked at in scenes containing clones using an eyetracker. Their results showed that heads and upper torsos are fixated significantly longer and more often when looking for appearance clones, while pelvis and lower torsos are found to be the most salient body parts when looking for motion clones. Even though these experiments help to better understand what influences clone detection, and how to make clones harder to spot, they do not answer the question of the minimum amount of variety necessary to avoid clone detection. This question was later investigated for motion clones by Pražák and O’Sullivan [2011], who demonstrated that cloning 3 unique motions (with desynchronisation for each character) was sufficient to create virtual crowds that were perceived as varied as when using a unique motion for each character. Although more directly applicable to our question, those results are limited to small crowds

of 8 to 24 characters and generic appearances, and it is therefore unclear whether they would generalise to larger crowds, as displaying more characters might for instance make clones easier to spot. Also, Hoyet et al. [2013] showed that the motions of some actors can be more distinctive than others, which may affect the perception of motion clones if a small number of distinctive motions are selected to animate a large number of characters.

While these work focus on how motion or appearance clones appear to the viewer, Shi et al. [2017] studied how variety in body shape can be used to create a cost-effective realistic crowd. An interesting point is raised about how distinctive the selected samples should be. Selecting highly distinctive shapes creates wider differences between varied characters but causes their repetition to be more easily spotted, while choosing more average shapes reduces overall variety but makes clones less apparent.

Creating variety in crowds has also been explored through how characters interact in a virtual crowd. For instance, Durupinar et al. [2011] proposed to create variations in the characters' trajectory by mapping simulation parameters to different personality traits, therefore leading to different interaction behaviours between characters. Focusing more on character full-body animations, Hoyet et al. [2016] demonstrated that adding secondary shoulder motions to characters passing at close distances can help to hide collisions between characters, while improving the overall realism of crowd scenes by introducing some variety in the displayed animations.

Virtual crowd perception and interactions have also been studied in Virtual Reality, to investigate how users interact when immersed in a virtual crowd. For instance, studies explored user eye-gaze behaviours in virtual crowds [Berton et al. 2020] or during dyadic interactions [Berton et al. 2019], speed adaptations to neighbouring agents when walking with virtual characters [Rio et al. 2018], following behaviours during evacuations [Ríos and Pelechano 2020], or the benefits of using haptic rendering of collisions while navigating through a dense virtual crowd [Berton et al. 2020].

*Synthesizing motion variations.* is also relevant to creating variety in crowds, as the overall goal is to generate new motion variations from several existing motions. This effectively creates additional motions that can enhance the perceived diversity of virtual crowd scenes, therefore these methods can be a cost effective and personalisable alternative to traditional motion capture. Most of these work do not target virtual crowds as their first application, but rather focus on adapting existing motions to new contexts and diversifying the range of motion of an actor without resorting to additional motion capture. Several methods [Min et al. 2010; Urtasun et al. 2004] use Principal Component Analysis to represent motions in a normalized, parameterizable space according to some characteristics (e.g., speed or step length), then generate new motions given a set of parameters. Another class of approaches is solely directed toward generating statistical variations of a set of source motions, without always providing controls on characteristics. These statistics-based models [Lau et al. 2009; Ma et al. 2010; Zhou et al. 2014] use a Bayesian Network to sample variations of the existing motions and blend them, while attempting to maintain coherence between the chosen samples. In the case of the model we chose to implement [Ma et al. 2010], both motion characteristics and random variations can be generated and parameterised.

### 3 EXPERIMENT

In this experiment, we are investigating the perception of motion clones in large-scale crowds (up to 1000 characters). More specifically, we are interested in answering the following questions. Can large-scale crowds with some degrees of motion uniformity be distinguished from the same crowds animated using a unique motion variation for each character? Is there a threshold to the relative number of motions used above which clones are usually undetected? Does this threshold depends on the total number of characters presented to the viewers?

To explore these questions, we conducted an experiment where users were presented with video pairs of virtual crowds, involving different numbers of characters (from 250 to 1000). In each video, characters were either animated with a unique synthesized motion, or with a relative subset of these motions. We hypothesised that

- H1. The lack of variety would be more easily noticed by viewers in virtual crowds using fewer number of unique motions (i.e., using more motion clones)
- H2. The lack of variety would be less noticed as the number of unique motions increased until a threshold, above which participants would not perceive differences anymore.
- H3. The small minimum number of unique motions (3) identified in previous studies would not hold for larger crowds, as motion clones would be significantly more present and easy to spot [McDonnell et al. 2008], and that some, potentially more distinctive, motion clones might also be more easily identified [Hoyet et al. 2013].

#### 3.1 Stimuli Creation

*Motion variations.* As it is not possible to acquire unique individual motions for such numbers of characters involved in large-scale crowds, we implemented the approach of Ma et al. [2010] to create sufficient statistical variations for our study. We chose this method over other existing methods described in Section 2, as it can generate large numbers of motion variations from a database containing a small number of examples. In short, the approach generates motions in two steps. First, for each desired motion to be generated, a set of *style parameter* is specified. For example the desired output motion is one of a female of average height, walking at average speed. From this set of input style parameters, *variation parameters* are computed for each limb using a Bayesian Network (BN). In the second step, both style and variation parameters are fed into a Kriging model that computes interpolation weights used to interpolate database motions together such as to reflect the characteristics given by the style and variation parameters in the resulting interpolated motion. In addition to these two fundamental steps, we applied a pre-processing step to normalize input motions in time and space using timewarping to synchronise footsteps [Witkin and Popovic 1995], as well as to retarget them to a common skeleton [Kulpa et al. 2005]. A post-processing step was also necessary to inverse the process and retarget the motions to either a male or female character, as well as to correct potential artifacts (e.g., footsliding).

Naturally, as the selected synthesis method is data-driven, both the BN and the Kriging model parameters are learnt from a database of existing motions. For our study, we used 156 motions captured



from 43 actors (20F, 23M) walking at their comfort speed. These animations were provided by the authors of [Hoyet et al. 2013], some of which were previously used for other experiments as well [Zibrek et al. 2020]. Their animations were obtained using a 19-camera Vicon optical tracking system, with 67 markers positioned on each actor’s body. We used 4 walking clips per actor, and although motions from the same actor are naturally less varied they still provide additional variety for the model to interpolate from.

As mentioned above, the motion synthesis model is driven by a set of style parameters, for which we chose to include actor gender and step length. Gender was chosen as a way to keep generated motions coherent with character appearance. Step length was chosen as it allowed further control on the distribution of walking speeds. It is to note that additional parameters could have been chosen to reflect physical features of the characters, like morphology or age. However, because we lacked sufficiently varied motions according to those additional parameters, we chose to limit the complexity of the model to the two aforementioned parameters.

Finally, we created 1000 motions using the described synthesis method, equally distributed in gender, and with step lengths randomly sampled from a normal distribution matching the database motions ( $\mu = 0.77\text{m}$ ,  $\sigma = 0.06\text{m}$ ). Variation parameters were then chosen randomly by the BN to synthesize each motion.

*Crowd trajectories.* For character trajectories, we chose to use a unidirectional crowd flow scenario, as illustrated in Figure 1. Simulations were created using the open-source software UMANS [van Toll et al. 2020], which implements various state-of-the-art algorithms, from which we chose to use the recent Dutra vision-based model [Dutra et al. 2017]. To realistically model the behaviour of characters in a crowd, we randomly assigned a preferred speed for each character using a normal distribution representing the distribution of speeds in the synthetic animations ( $\mu = 1.44\text{m/s}$ ,  $\sigma = 0.13\text{m/s}$ , clipped to  $[1.0, 1.88]\text{m/s}$ ). Using a range of speeds for the characters was also beneficial to generate more varied character animations by making better use of the set of interpolated motions by the motion synthesis model. It is also important to mention that for each trajectory the corresponding animation started at a random frame of the walk cycle, to desynchronize all the animations (including motion clones) and avoiding the perturbing effect of having all characters walking in-step.

*Character models.* Because appearance clones are more easily noticed than motion clones [McDonnell et al. 2008], sufficiently varied character appearances should be used to avoid disturbing the perception of motion variety. We used a set of 40 characters (20 males and 20 females), each featuring 6 texture variations for a total of 240 appearance variations.

*Viewpoint.* We chose to use a canonical aerial viewpoint for the generation of our stimuli, similar to those used in previous studies on the perception of virtual crowds [Cabreró Daniel et al. 2021; Hoyet et al. 2016; Kulpa et al. 2011; Pražák and O’Sullivan 2011], with a low enough field of view to allow picking out details almost evenly throughout the crowd. Because both scenes needed to be shown on the same screen, they were recorded at a half 16:9 aspect ratio. Recordings were made at 1920x2160 pixels at 60Hz and downscaled to 960x1080 pixels to obtain better details than with

**Table 1: Number of unique motions used for each condition. \*number of male and female motions were counter-balanced over the experiment in these conditions.**

		Crowd Size		
		250	500	1000
<b>Gold Standard</b>	<b>1:1</b>	125F,125M	250F,250M	500F,500M
	<b>1:2</b>	63F,62M*	125F,125M	250F,250M
<b>Variation Ratio</b>	<b>1:5</b>	25F,25M	50F,50M	100F,100M
	<b>1:10</b>	13F,12M*	25F,25M	50F,50M
	<b>1:25</b>	5F,5M	10F,10M	20F,20M
	<b>1</b>	1F,1M	1F,1M	1F,1M

anti-aliasing. Each resulting videos was 28 seconds long. Examples of the stimuli are presented in the companion video.

### 3.2 Experiment Design

*Protocol.* Participants were presented with video pairs of crowd scenes. On one video of each pair, every character was animated with a unique synthesized motion, while on the other video the characters’ animations were a subset of these motions. The participants’ task was to select the video which they considered had the most unique animations. To generate these videos, we used the 1000 synthetic motions created using the approach detailed above.

As we were interested in the relative number of motion variations presented to participants, additionally to gold standard videos displaying a unique motion for each character we created videos with 5 different levels of *Variation Ratio*: 1:2, 1:5, 1:10, 1:25, 1. 1:N means that one motion is used for N different characters, while the 1 value corresponds to a single motion used for all the characters of the same gender. It is important to mention that in all cases, the number of varied motions is split between male and female characters, as there are an equal number of each. For the specific 1 value, we chose to use an absolute number rather than a ratio in order to represent the worse case scenario where only one motion example per gender is used to animate a whole virtual crowd. Also, as using distinctive motions might highly affect the perception of motion clones in such a case, we chose to focus on the most difficult situation where none of the motions are visually outstanding. For the 1 condition, we therefore randomly used one of the three least distinctive male and female captured motions from [Hoyet et al. 2013], as it would be normal to be demanding about the animations available when only a few examples are required.

As we were also interested in exploring whether the number of characters present in the crowd might affect the perception of motion clones, we generated crowd scenes with 3 different *Crowd Size*, displaying in total 250, 500 or 1000 characters. The higher the size of the crowd, the larger the set of different motions used, and the more varied the scene might seem. However, the scene is also more cluttered, which might make it more difficult to spot clones. The exact number of motions used for each condition is summarized in Table 1.

In total, participants were presented with 45 video-pair comparisons: 5 *Variation Ratio*  $\times$  3 *Crowd Size*  $\times$  3 *Repetitions*. In each condition, the video was presented side-by-side with the golden

standard video using a unique synthetic motion for each character. The golden standard video was randomly presented on the left or right side of the screen, and the presentation order of the conditions was also randomized. Both videos in a pair played simultaneously, were 28s long, and could be visualised twice. After that, we asked participants to select the scenario they perceived to have the most unique animations, if they had not already provided their answer. The study was conducted online, however videos were all loaded locally before the experiment began, ensuring that they would play without interruption at the original bitrate and resolution. They were presented to participants in full screen.

**Participants.** Twenty participants (12F, 8M; age: avg=29.9, min=18, max=66) performed the experiment online. They were recruited using internal mailing lists, and did not receive any compensation. They were all naive to the purpose of the experiment, and gave their informed consent prior to the experiment. Data collected from participants were fully anonymised. The study conformed to the declaration of Helsinki.

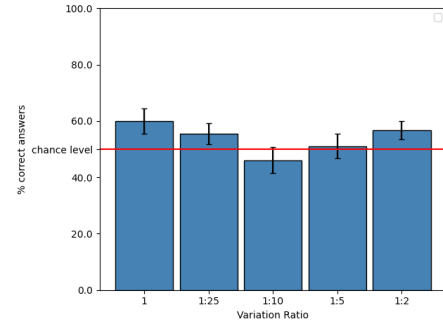
### 3.3 Analysis

For each participant, we computed the percentage of correct answers for each *Variation Ratio* and *Crowd Size* condition across the three repetitions, where an answer was considered to be correct if participants selected the gold standard video (i.e., displaying a unique synthetic motion for each character). We assessed normality of the data using a Shapiro-Wilk normality test. As residuals were found to be non-normally distributed, and the participants' answers are discrete, we transformed this non-parametric data using the Aligned Rank Transform procedure [Wobbrock et al. 2011]. With this method, the data is aligned and ranked before being tested with a traditional Analysis of Variance (ANOVA) to analyze interaction effects. We then performed a two-way Repeated Measures ANOVA with within-subject factors *Variation Ratio* (1:2, 1:5, 1:10, 1:25, 1) and *Crowd Size* (250, 500, 1000) on the percentage of participants' correct answers. Bonferroni post-hoc comparisons were performed when a significant main or interaction effect was found. We also performed a t-test on the percentage of correct answers against the 50% chance level threshold, to assess whether participants' answers were significantly above chance.

## 4 RESULTS

We found a main effect of the *Variation Ratio* ( $F_{4,266} = 2.6804, p = 0.032109$ ) on the percentage of correct answers, suggesting that some amounts of motion variations were distinguished significantly more often by participants (Figure 2). However, post-hoc analysis showed a significant difference only between 1 and 1:10 *Variation ratios* ( $p = 0.0318$ ), which does not provide any meaningful insight, and is likely due to chance. No other comparisons were significant, even with the 1:25 ratio that is in between 1 and 1:10. No other main or interaction effect was found: crowd size did not influence the detection of the gold standard, regardless of the amount of variation present in the other scene.

As mentioned previously, we also conducted single t-tests to assess whether the percentage of correct answers was significantly above the 50% chance level threshold. Results showed that only the crowds animated with a single motion per gender (1 *Variation Ratio*)



**Figure 2: Percentage of correct answers for each *Variation Ratio*. The chance level (red) indicates a statistically random choice. Error bars represent the Standard Error of the Mean.**

were recognised often enough by participants to be significantly above the 50% chance level ( $F = 2.213882, p = 0.039265$ ). For higher *Variation Ratios*, participants did not chose the gold standard frequently enough to conclude of a significant perceptual difference between the two scenes. These results suggest that viewers can identify crowds animated with 2 motions only (one per gender), and that introducing a few more varied animations is perceptually equivalent to animating each character with its own unique motion.

## 5 DISCUSSION

Because of the cost of animating human crowds, as well as budget constraints, a small number of motions is often used to animate a larger set of characters. In this study, we were interested in whether results from previous studies exploring how many unique motions are required for motion clones to be imperceptible in small groups of characters (up to 24 characters) would remain valid for large-scale crowds (up to 1000 characters).

Our results showed that only virtual crowds animated with 2 unique motions (one male and one female) were perceived as significantly different from the gold standard, invalidating **H1** and **H2**. Participants could not therefore identify the other conditions as having fewer unique motions than the gold standard, regardless of crowd size (also invalidating **H3**). First, it is both surprising and interesting to have obtained such a low accuracy for the large-scale crowd conditions involving small numbers of unique motions, especially for the condition involving only one unique female and male motion (60% accuracy), showing the overall difficulty of the task. Difficulty ratings provided by participants after our study further emphasize this fact, as 18 out of 20 of our participants found the task Difficult or Very Difficult. This level of accuracy is however of a similar magnitude than the same condition in previous work involving small-scale crowds ( $\approx 63\%$  accuracy [Pražák and O'Sullivan 2011]). It is also interesting to notice that participants performed similarly in both studies, despite differences in the number of characters, or from the fact that we deliberately chose to use validated non-distinctive motions in our experiment (as it would be normal to be demanding about animation quality when using only a few examples). This result suggests that extremely low numbers of unique motions could be considered even for large-scale crowds.

It should also be mentioned that the small participant sample size of our ( $N=20$ ) and their ( $N=12$ ) study might not reflect the accuracy of the general population, and that a larger number of participants might provide stronger evidence that differences in accuracy for higher *Variation Ratios* in our results are not due to chance.

As mentioned above, we also found that participants could not significantly discriminate between the gold standard and crowds using variation ratios of one motion for 25 characters or less. This result does not support our hypothesis that previous result of using only 3 motions might not hold for large-scale crowds, as this condition corresponded to using between 10 and 40 motions in our experiment depending on the crowd size. Also, similarly to previous work [Pražák and O'Sullivan 2011], our results showed that the perception of motion clones do not seem to be influenced by the size of the crowd presented to participants. This parallel seems to confirm that crowd size does not play a significant role in motion clone detection, and that their perception in large-scale crowds follow the same mechanisms as in small-scale ones.

It is also possible that participants' overall low accuracy could be related to limitations in our study. In particular, as it is not possible to acquire unique motion capture examples for each character of large-scale crowds, we relied on a state-of-the-art motion variation approach to synthesize unique motion variations. While most of these motions were validated to show a relative range of distinctiveness [Hoyet et al. 2013], their choice and number might have influenced how varied the synthetic motions were. For instance, if synthetic motions seemed too similar to viewers, the various crowd stimuli might have been perceived as uniform, and differences in variation levels more difficult to spot. As the synthesis method is based on motion blending, the number of actors and motion clips in the database, as well as their overall distinctiveness, set a limit to the truly varied combinations that can be obtained. Another crucial parameter is the quality and naturalness of synthetic motions, as artifacts and unnatural gaits could draw viewers' attention and help them spot multiple characters exhibiting the same behaviour. While the Bayesian Network ensured that variation parameters of all body parts comparatively matched the database motions, we noticed while visualizing individual synthetic motions some small visual artifacts where the motion of a body part seemed to be mismatched with the rest of the body. This proved to be a limitation of the model, and despite addressing most of its consequences, some motions still displayed uneven swings of the arms. While this was visible on individual motions alone, it is unclear whether it could have also affected viewers' responses as well, and a more thorough evaluation of the motions generated, both in terms of diversity in variations and naturalness, would help to further understand the actual variations that can be synthesised in the future.

Chosen character speeds and trajectories could also have affected our results. Although 3 different sets of speeds and trajectories were used for each crowd size, they all shared the same flow direction, speed distribution, and avoidance algorithm. Varying speeds means that quicker characters try to pass slower ones, creating clutters and gaps. Those kind of irregularities, could have diverted the attention of viewers from the actual character motions. A unidirectional flow direction also means that characters are mostly seen from the same angle, and coupled with the choice of a low field of view our stimuli presented almost all characters from the same angle.

However, observing characters from a remote viewpoint might have diminished participants' ability to perceive the details of their motion and prevented them from spotting motion clones. Exploring the effect of point of view on crowd perception therefore seems to be important for future work, especially as it was recently found that an eye-level point of view seems to enable users to more accurately perceive visual artifacts in character trajectories, possibly because of being closer to the one we are used to experience as humans [Cabrero Daniel et al. 2021].

It also seems important to point out that the *Crowd Size* factor corresponded to the total number of characters showed during the entire stimuli. We estimated the actual number of characters on screen at any moment to be about a fifth (55, 110 and 220 characters respectively). This means that there were fewer motion clones simultaneously present on screen than if all the characters had been always present, which might also have made them more difficult to spot. Moreover, participants likely had to rely on memory to spot other instances of previously seen motion clones, since fewer direct comparisons were possible, unlike in previous studies.

Finally, because of the pandemic the experiment was conducted online, and a number of experimental factors that were difficult to control could also have affected results. E.g., conditions like screen size and resolution could not be normalized across participants. In particular, only 11 participants provided the size of the screen they performed the experiment on (avg=21.7, max=55, min=14). Nine participants used screens with resolution between 720p and 1080p, nine others used screens at 1080p resolution, the two remaining participants had screens of 1440p and 2160p resolution. This also demonstrates the variety of displays that can be potentially used in graphics applications nowadays, and the corresponding range of graphics quality required by such applications.

## 6 CONCLUSION

In this study, we investigated the perception of different levels of motion variations in large-scale virtual human crowds. A perceptual experiment was conducted to record participants' accuracy at distinguishing fully varied crowds from less varied ones. Our findings suggest that in large-scale crowds even low levels of motion variety (1 motion for 25 characters) do not seem to be perceivable by users in comparison to fully varied crowds. In our study, only crowds animated with 2 motions (1 per gender) tended to be noticed as less varied by viewers. Participants' overall lack of accuracy, even on less varied stimuli, seems to validate in the case of large-scale crowds observations made in previous studies on small-scale crowds, that motion variations in crowds are quite subtle and motion clones difficult to pick out. Those empirical results could guide creators and developers decide on the resources they should allocate to motion variety in their crowd application based on a formal user study rather than on experience or general feeling. With this paper, we have improved the comprehension of motion perception in crowds, and we believe that further studies will permit more realistic and efficient crowd applications.



## ACKNOWLEDGMENTS

We wish to thank all the reviewers for their comments, and the participants in our experiments. This work was funded by the French ANR as part of the JCJC Per<sup>2</sup> project (ANR-18-CE33-0013).

## REFERENCES

- A. Beacco, N. Pelechano, and C. Andújar. 2016. A Survey of Real-Time Crowd Rendering. *Computer Graphics Forum* 35, 8 (2016), 32–50. <https://doi.org/10.1111/cgf.12774>
- T Beardsworth and T Buckner. 1981. The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society* 18, 1 (1981), 19–22.
- Florian Berton, Fabien Grzeskowiak, Alexandre Bonneau, Alberto Jovane, Marco Aggravi, Ludovic Hoyet, Anne-Helene Olivier, Claudio Pacchierotti, and Julien Pettre. 2020. Crowd Navigation in VR: exploring haptic rendering of collisions. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. <https://doi.org/10.1109/TVCG.2020.3041341>
- Florian Berton, Anne-Hélène Olivier, Julien Bruneau, Ludovic Hoyet, and Julien Pettre. 2019. Studying Gaze Behaviour during Collision Avoidance with a Virtual Walker: Influence of the Virtual Reality Setup. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 717–725. <https://doi.org/10.1109/VR.2019.8798204>
- B. Cabrero Daniel, R. Marques, L. Hoyet, J. Pettré, and J. Blat. 2021. A Perceptually-Validated Metric for Crowd Trajectory Quality Evaluation.
- Thierry Chaminade, Jessica Hodgins, and Mitsuo Kawato. 2007. Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience* 2, 3 (05 2007), 206–216.
- James E. Cutting and Lynn T. Kozlowski. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society* 9, 5 (1977), 353–356.
- Simon Dobbyn, John Hamill, Keith O'Connor, and Carol O'Sullivan. 2005. Geopostors: A Real-Time Geometry/Impostor Crowd Rendering System. *ACM Trans. Graph.* 24, 3 (July 2005), 933. <https://doi.org/10.1145/1073204.1073290>
- Funda Durupinar, Nuria Pelechano, Jan Allbeck, Ugur Gudukbay, and Norman I. Badler. 2011. How the Ocean Personality Model Affects the Perception of Crowds. *IEEE Computer Graphics and Applications* 31, 3 (2011), 22–31. <https://doi.org/10.1109/MCG.2009.105>
- T. B. Dutra, R. Marques, J.B. Cavalcante-Neto, C. A. Vidal, and J. Pettré. 2017. Gradient-based steering for vision-based crowd simulation algorithms. *Computer Graphics Forum* 36, 2 (2017), 337–348. <https://doi.org/10.1111/cgf.13130>
- Jessica Hodgins, Sophie Jörg, Carol O'Sullivan, Sang Il Park, and Moshe Mahler. 2010. The Saliency of Anomalies in Animated Human Characters. *ACM Trans. Appl. Percept.* 7, 4, Article 22 (2010). <https://doi.org/10.1145/1823738.1823740>
- Ludovic Hoyet, Rachel McDonnell, and Carol O'Sullivan. 2012. Push It Real: Perceiving Causality in Virtual Interactions. *ACM Trans. Graph.* 31, 4, Article 90 (July 2012). <https://doi.org/10.1145/2185520.2185586>
- Ludovic Hoyet, Anne-Helene Olivier, Richard Kulpa, and Julien Pettré. 2016. Perceptual Effect of Shoulder Motions on Crowd Animations. *ACM Trans. Graph.* 35, 4, Article 53 (July 2016). <https://doi.org/10.1145/2897824.2925931>
- Ludovic Hoyet, Kenneth Ryall, Katja Zibrek, Hwangpil Park, Jehée Lee, Jessica Hodgins, and Carol O'Sullivan. 2013. Evaluating the Distinctiveness and Attractiveness of Human Motions on Realistic Virtual Bodies. *ACM Trans. Graph.* 32, 6, Article 204 (Nov. 2013). <https://doi.org/10.1145/2508363.2508367>
- Gunnar Johansson. 1973. Visual perception of biological motion and model for its analysis. *Perception & Psychophysics* 14, 2 (1973), 201–211.
- Lynn T. Kozlowski and James E. Cutting. 1977. Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics* 21, 6 (1977), 575–580.
- Richard Kulpa, Franck Multon, and Bruno Arnaldi. 2005. Morphology-independent representation of motions for interactive human-like animation. *Computer Graphics Forum* 24, 3 (2005), 343–351. <https://doi.org/10.1111/j.1467-8659.2005.00859.x>
- Richard Kulpa, Anne-Hélène Olivier, Jan Ondřej, and Julien Pettré. 2011. Imperceptible Relaxation of Collision Avoidance Constraints in Virtual Crowds. *ACM Trans. Graph.* 30, 6 (Dec. 2011). <https://doi.org/10.1145/2070781.2024172>
- Manfred Lau, Ziv Bar-Joseph, and James Kuffner. 2009. Modeling spatial and temporal variation in motion data. In *ACM Transactions on Graphics (TOG)*, Vol. 28. ACM, 171.
- Wanli Ma, Shihong Xia, Jessica K Hodgins, Xiao Yang, Chunpeng Li, and Zhaoqi Wang. 2010. Modeling style and variation in human motion. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '10)*. 21–30.
- Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona N. Newell, and Carol O'Sullivan. 2009a. Investigating the Role of Body Shape on the Perception of Emotion. *ACM Trans. Appl. Percept.* 6, 3, Article 14 (Sept. 2009), 11 pages.
- Rachel McDonnell, Michéal Larkin, Simon Dobbyn, Steven Collins, and Carol O'Sullivan. 2008. Clone Attack! Perception of Crowd Variety. *ACM Trans. Graph.* 27, 3 (Aug. 2008), 1–8. <https://doi.org/10.1145/1360612.1360625>
- Rachel McDonnell, Michéal Larkin, Benjamin Hernández, Isaac Rudomin, and Carol O'Sullivan. 2009b. Eye-Catching Crowds: Saliency Based Selective Variation. *ACM Trans. Graph.* 28, 3, Article 55 (July 2009). <https://doi.org/10.1145/1531326.1531361>
- Jianyuan Min, Huajun Liu, and Jinxiang Chai. 2010. Synthesis and Editing of Personalized Stylistic Human Motion. In *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D '10)*. 39–46. <https://doi.org/10.1145/1730804.1730811>
- Carol O'Sullivan. 2009. Variety Is the Spice of (Virtual) Life. In *Proceedings of the 2nd International Workshop on Motion in Games (MIG '09)*. 84–93. [https://doi.org/10.1007/978-3-642-10347-6\\_8](https://doi.org/10.1007/978-3-642-10347-6_8)
- Martin Pražák, Ludovic Hoyet, and Carol O'Sullivan. 2011. Perceptual Evaluation of Footskate Cleanup. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '11)*. 287–294. <https://doi.org/10.1145/2019406.2019444>
- Martin Pražák and Carol O'Sullivan. 2011. Perceiving Human Motion Variety. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV '11)*. 87–92. <https://doi.org/10.1145/2077451.2077468>
- Kevin W Rio, Gregory C Dachner, and William H Warren. 2018. Local interactions underlying collective motion in human crowds. *Proceedings of the Royal Society B: Biological Sciences* 285, 1878 (2018), 20180611.
- A. Ríos and N. Pelechano. 2020. Follower behavior under stress in immersive VR. *Virtual Reality* 24 (2020), 683–694. <https://doi.org/10.1007/s10055-020-00428-8>
- Yinxuan Shi, Jan Ondřej, He Wang, and Carol O'Sullivan. 2017. Shape up! Perception based body shape variation for data-driven crowds. In *2017 IEEE Virtual Humans and Crowds for Immersive Environments (VHCIE)*. <https://doi.org/10.1109/VHCIE.2017.7935623>
- Nikolaus. F. Troje. 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision* 2 (2002), 371–387.
- Raquel Urtasun, Pascal Gdard, Ronan Boulic, Daniel Thalmann, and Pascal Fua. 2004. Style-based motion synthesis. In *Computer Graphics Forum*, Vol. 23. Wiley Online Library, 799–812.
- Wouter van Toll, Fabien Grzeskowiak, Axel López Gandía, Javad Amirian, Florian Berton, Julien Bruneau, Beatriz Cabrero Daniel, Alberto Jovane, and Julien Pettré. 2020. Generalized Microscopic Crowd Simulation Using Costs in Velocity Space. In *Symposium on Interactive 3D Graphics and Games (I3D '20)*. Article 6. <https://doi.org/10.1145/3384382.3384532>
- W. van Toll and J. Pettré. 2021. Algorithms for Microscopic Crowd Simulation: Advancements in the 2010s. *Computer Graphics Forum* 40, 2 (2021), 731–754. <https://doi.org/10.1111/cgf.142664>
- Michele Vicovaro, Ludovic Hoyet, Luigi Burigana, and Carol O'sullivan. 2014. Perceptual Evaluation of Motion Editing for Realistic Throwing Animations. *ACM Trans. Appl. Percept.* 11, 2, Article 10 (2014). <https://doi.org/10.1145/2617916>
- Andrew Witkin and Zoran Popovic. 1995. Motion Warping. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*. 105–108. <https://doi.org/10.1145/218380.218422>
- Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.
- Liuyang Zhou, Lifeng Shang, Hubert PH Shum, and Howard Leung. 2014. Human motion variation synthesis with multivariate Gaussian processes. *Computer Animation and Virtual Worlds* 25, 3-4 (2014), 301–309.
- Katja Zibrek, Ludovic Hoyet, Kerstin Ruhland, and Rachel McDonnell. 2015. Exploring the Effect of Motion Type and Emotions on the Perception of Gender in Virtual Humans. *ACM Trans. Appl. Percept.* 12, 3, Article 11 (July 2015), 20 pages.
- Katja Zibrek, Benjamin Niay, Anne-Hélène Olivier, Ludovic Hoyet, Julien Pettre, and Rachel McDonnell. 2020. The Effect of Gender and Attractiveness of Motion on Proximity in Virtual Reality. *ACM Trans. Appl. Percept.* 17, 4, Article 14 (Nov. 2020). <https://doi.org/10.1145/3419985>