EUROGRAPHICS 2022 / R. Chaine and M. H. Kim
(Guest Editors)

Volume 41 (2022), Number 2

Interaction Fields: Intuitive Sketch-based Steering Behaviors for Crowd Simulation

Supplementary Material

A. Colas[1], W. van Toll[1,2], K. Zibrek[1], L. Hoyet[1], A.-H. Olivier[1], J. Pettré[1]

[1]Univ Rennes, Inria, CNRS, IRISA, France [2] Breda University of Applied Sciences, The Netherlands

In this supplementary material, we further detail the user study conducted to evaluate IFs. It is important to point out that our general goal was not to evaluate the actual design of the GUI, but rather the ease and ability of non-expert users to learn and independently leverage the functionalities of IFs to design scenarios involving interactions of increasingly complexity.

1. Pilot study

Prior to the actual study, we conducted an online pilot with 6 participants that were familiar with computer animation but not with interaction fields. Our goal was to collect initial feedback through an online form about the difficulty and duration of the initial training and tasks we intended to include in the experiment. The purpose was not to evaluate the actual design of the GUI, but rather improve upon or include specific GUI functionalities based on user feedback. Following the pilot study, we identified a need for reorganizing the experiment to both shorten duration and facilitate the learning process, in particular in terms of the number and order of the tasks to perform.

2. Protocol

For the subsequent user study, participants were invited to take part in the experiment at our research institute. Upon arrival, they were first asked to read and sign a consent form. The experiment itself used two 24-inch screens and a mouse and keyboard, as shown in Figure 1. On the left screen, participants saw the IF editor GUI in which they could draw interaction fields. On the right screen, they saw the resulting simulation, in which the drawn IFs were emitted by a source (in red) and received by another agent (in yellow). Participants were always allowed to refine their IF sketches on the left and play a new simulation on the right, until they were satisfied with their result and marked a task as completed.

Instructions were displayed on the computer screens as the experiment progressed. The experiment consisted of an initial exploration phase, followed by five scenarios of increasing complexity.

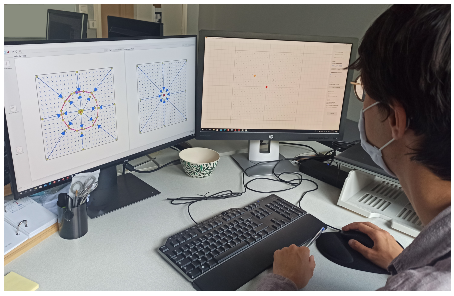


Figure 1: *Photo of a user conducting our user study.*

The total duration of the experiment depended strongly on the participant, but did not exceed 2 hours.

Exploration. The aim of the exploration phase was to give participants a first glance of IFs. They were given a short video-guided introduction of its uses and applications, and the opportunity to freely explore our IF tool for ten minutes. They were told to play with the GUI, to test simple drawings, and to visualize results in the simulation window. This stage helped us gauge the initial learnability of the GUI and the users' first impression. Participants were also allowed to interact with the experimenter and to ask questions.

Scenarios. After the exploration phase, participants were asked to draw IFs for specific agent behaviors in a sequence of scenarios. Each scenario started with a training example, followed by one to three evaluation tasks. The seven evaluation tasks in total are summarized in Table 1.

- **Training tasks.** Each scenario started with a training example covering a specific concept of IFs (e.g., controlling velocity, controlling orientation, creating parametric IFs.). For each concept, participants were provided precise instructions to use the new functionality effectively. They were first instructed to attempt to draw the matching field by themselves, and then to follow a

Scenario	Task	Instructions	Expert rating criteria	Expert ratings	Inter-rater reliability Fleiss Kappa
S1	Task 1	The goal is to draw a velocity IF that makes the yellow agent move and stop 5 meters away from the red agent.	The yellow agent should: • be repulsed when too close to the red agent • be attracted when too far from the red agent • stop at a 5m distance • keep respecting this distance	9.13 ± 0.52	0.34 : Fair
S2	Task 2	The goal is to draw a velocity IF that makes the yellow agent circle counter-clockwise around the red agent at a distance of 5 meters, combined with an orientation IF that makes the yellow agent look at the red agent.	The yellow agent should: • repulsed when close to the red agent • attracted when far from the red agent • turn counter-clockwise • respect a 5m distance • look at the red agent	9.67 ± 0.52	0.87 : Very good
S3	Task 3	In this task, the agents play 'red light, green light'. The goal is to draw a velocity IF that makes the yellow agents move toward the red agent when the red agent is looking away, and stand still when the red agent looks at them.	The yellow agent should: • stop when the red agent is looking • move towards the red agent when it is not looking	9.77 ± 0.00	0.96 : Very good
	Task 4	The goal is to draw a velocity IF and an orientation IF that make the yellow agent • move and look towards the red agent when the red agent is looking away, • stand still while turning its back to the red agent when the red agent is looking towards them.	The yellow agent should: • stop when the red agent is looking • move towards the red agent when it is not looking • look away from the red agent when it is looking • look towards the red agent when it is not looking	9.58 ± 0.35	0.81 : Very good
	Task 5	The goal is to draw a velocity IF that makes the yellow agent stand and move on the right side of the red agent, 2 meters away from it.	The yellow agent should: • go to the right side of the red agent • respect a 2m distance to the red agent • maintain its relative position	9.44 ± 0.13	0.72 : Good
S4	Task 6	The goal is to draw a parametric velocity IF that makes the yellow agent stand and move on the right side of the red agent. Unlike in task 5, the yellow agent must stay close when the red agent is moving slowly (1 meter way), and remain further away when the red agent is moving faster (5 meters away).	The yellow agent should: • go to the right side of the red agent • move further when speed is high • move closer when speed is low • respect 5m distance at high speed • respect 1m distance at low speed	9.51 ± 0.79	0.62 : Good
S5	Task 7	The goal is to draw a parametric velocity IF that makes the yellow agent move behind the obstacle to hide from the blue agent, combined with an orientation IF that makes the yellow agent look at the blue agent. The obstacle is the source of the velocity IF. The blue agent is the source of the orientation IF.	The yellow agent should: • always go to the opposite side of the obstacle • look towards the blue agent	10.00 ± 0.00	1.00 : Very good

Table 1: *Descriptions of the five scenarios and seven evaluation tasks in the user study. Columns show the overall goal of a task (which the participants received as instructions), the agent behavior criteria used during the expert evaluation, and the resulting expert grade (averaged over all users and experts combined).*

video tutorial that showed an approved way of designing the expected field. At the end of each training, participants were asked to answer two questions on a 7-point Likert scale: "I understood the concept of this training" and "I am satisfied with the ease of completing the training".

- **Evaluation tasks.** After each training, participants performed one or more evaluation tasks, to evaluate their ability to apply the functionalities learned in the previous training. They were only provided with written instructions describing the task to achieve (see Table 1) and a video showing the expected agent behavior (but not the expect IFs). While performing the evaluation tasks, participants were not allowed to ask questions to the experimenter. They were instructed to stop when they felt that their simulation was similar enough to the expected result presented in the video. At the end of each task, participants were asked to answer the following question on a 7-point Likert scale: "I am satisfied with the ease of completing this task" and "I am satisfied with the end result".

Final questionnaire. At the end of the experiment, participants were asked to answer a general questionnaire about the usability of the IF editor. This questionnaire is based on the System Usability Scale Questionnaire (SUS) [Bro96], which is commonly used to evaluate the perceived usability of commercial tools. However, since our current tool is still in development and was not designed for a commercial use at this stage, we decided to create our own questionnaire, see Table 2. We used 5 questions from the SUS questionnaire (see questions 2-6 in Table 2) and included two questions about error management (questions 7 and 8 in Table 2) and significance of code (questions 9 and 10 in Table 2) from [AE*16]. An additional question was added at the beginning of the questionnaire referring to the training for the system (question 1).

3. Results

Participants. 22 participants (7 women, 15 men; age: 28.4±8.0, min: 22, max: 62), volunteered for the study. They were all naive to the IF sketching tool, but had some knowledge of 3D animation or crowd simulation. They were recruited via internal mailing lists amongst students and staff. Participants gave their written and informed consent prior to the experiment. The study conformed to the declaration of Helsinki, and was reviewed and accepted by our Institutional Research Ethics Board (left blind for review).

To evaluate the ease and ability of non-expert users to learn and independently use IFs to design specific scenarios of interactions, we analysed both self-reported user experience with the sketch tool and expert evaluation of participants' simulation results.

Self-reported user experience. To evaluate participants' experience with each scenario, they answered two questions on a 7-point Likert scale after each training and evaluation task. These questions were chosen to assess their understanding of the concepts presented, the ease with which they performed the tasks, and their satisfaction with the agent behavior they designed.

The answers (summarized in Figure 2(a) and Figure 2(b)) show that participants found IFs easy to learn and to use, and that they were very satisfied with the results they could produce. In particular, participants understood each concept of the training, and they managed to apply these concepts easily in both the training and evaluation tasks. However, we noticed a tendency for slightly lower ratings for Task 3, which was the first task to introduce the concepts of the relativity and orientation of a moving source, while simultaneously requiring to define a large zero area. We believe that these two concepts might have been relatively hard to process at the same time. Nevertheless, the results show that participants could design the requested behaviors to their own satisfaction.

Response time. To evaluate how quickly participants were able to design agent behaviors, we recorded the amount of time spent by participants drawing IFs for each task. The timer of each task started with the first input to the sketch and ended with the last without interruption. Participants took on average 3.09 minutes to realize a task. Figure 2(c) shows that the majority of participants could quickly sketch their field to their satisfaction. However, some participants took a longer time to adjust, as indicated by a high standard deviation for some tasks. The inter-individual variability of the time stays acceptable with a maximum of 16 minutes.

Visual interpretation of results. Of course, not all participants drew the exact same interaction fields. Figure 3 gives a visual impression of the 'average' IF that participants drew for each individual task, as well as the variation among participants. It also shows the resulting agent trajectories for different participants in different colors. These results are also shown in motion in a dedicated supplementary video. While the IFs and trajectories vary among participants, the overall behavior remains visually comparable to what they were instructed to design.

Figures 3 (a) and (b) show a relatively high variance in the grid cells around a 5m distance from the source (where agents were expected to stop in these two tasks). This suggests that participants could not sketch the stopping distance constraints very precisely. This is explicable because of our grid-based IF representation: zero areas are approximated by empty cells, and not all participants selected the exact same cells to empty.

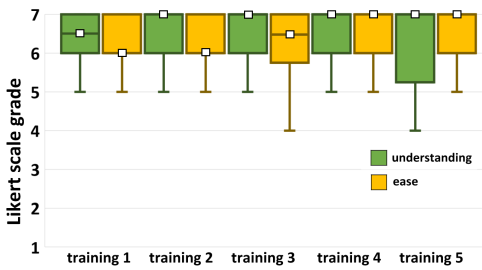
Some visual differences in the drawn fields can also be explained by the fact that participants interpreted some task instructions (e.g. words such as "towards" or "in front") in different ways. Such differences can be observed in Figure 3 (c) and (d), where some participants erased all the cells below the source's position while others tried to mimic a field of view and only erased part of them. Those differences were however not visible during the simulation because the trajectory of the source and the initial positions of the responding agents were fixed.

Expert evaluation of results. To further evaluate participants' simulation results, 3 experts amongst the authors rated the agent behavior that resulted from the IFs of each participant. Prior to the experiment, we defined a list of evaluation criteria per task, describing objectively the behavior that should be captured by the IFs of that task. These criteria are given in Table 1. Participants did not see these lists; they were expected to infer the requirements from their overall task instructions.

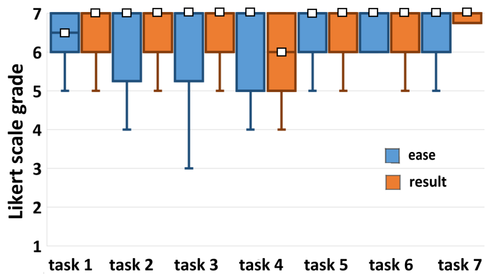
After the user study, each expert checked independently (i.e. without communicating with the other experts) which of these evaluation criteria were met for each individual result (i.e. for each participant's output in each task). Per task, we converted this expert evaluation to a rating on a scale of 0 to 10, based on the average number of satisfied criteria over all participants and experts. The resulting ratings per task are shown in Table 1 as well. As this table shows, the overall scores were very high, indicating that almost all users were capable of drawing IFs that met all behavioral criteria. To quantify the reliability of these scores, we performed an inter-rater reliability test per task using Fleiss' Kappa. The average overall score was 0.76, which is interpreted as good reliability. The scores per task are shown in the last column of Table 1.

We acknowledge that the 'correctness of agent behavior' is scenario-specific and difficult to quantify, and that a review with external experts would be considered as even more reliable. Still, we are confident that this expert evaluation is meaningful: in combination with the purely visual results discussed before, it serves the purpose of assessing whether users understood the instructions.

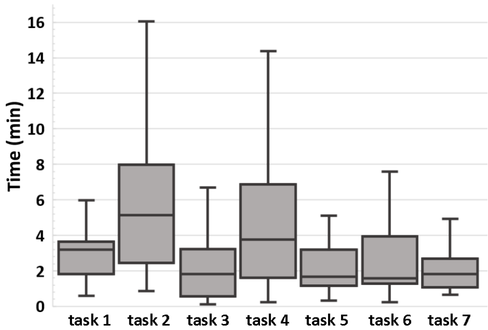
IF usability questionnaire. To assess participants' general satisfaction with our IF editor, we presented them with a usability questionnaire of which the questions and results are given in Table 2. To compute an overall usability score, we first inverted the negative items scores (8 – value), and then computed an average value over



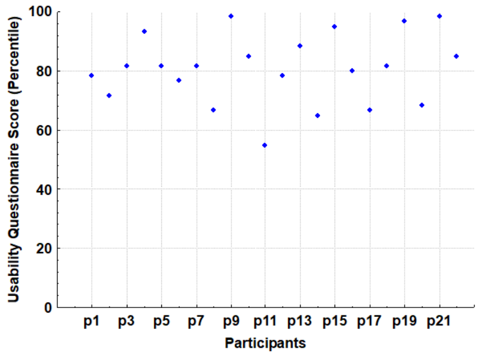
(a) User satisfaction (Training)



(b) User satisfaction (Evaluation tasks)



(c) Task completion time



(d) Usability scores

Figure 2: *(a) Boxplots showing the median ratings, interquartile ranges, and maximum/minimum ratings (outliers excluded) of the understanding of the training (green boxplots) and the ease of completing the training (yellow boxplots) for the 5 training tasks. (b) Average ratings and ranges for the 7 tasks participants completed independently. The blue boxplots represent participants' ease of completing the task and orange their satisfaction with the final result. (c) Boxplots representing the time to complete each task. (d) The final usability scores (in percentiles) for each participant.*

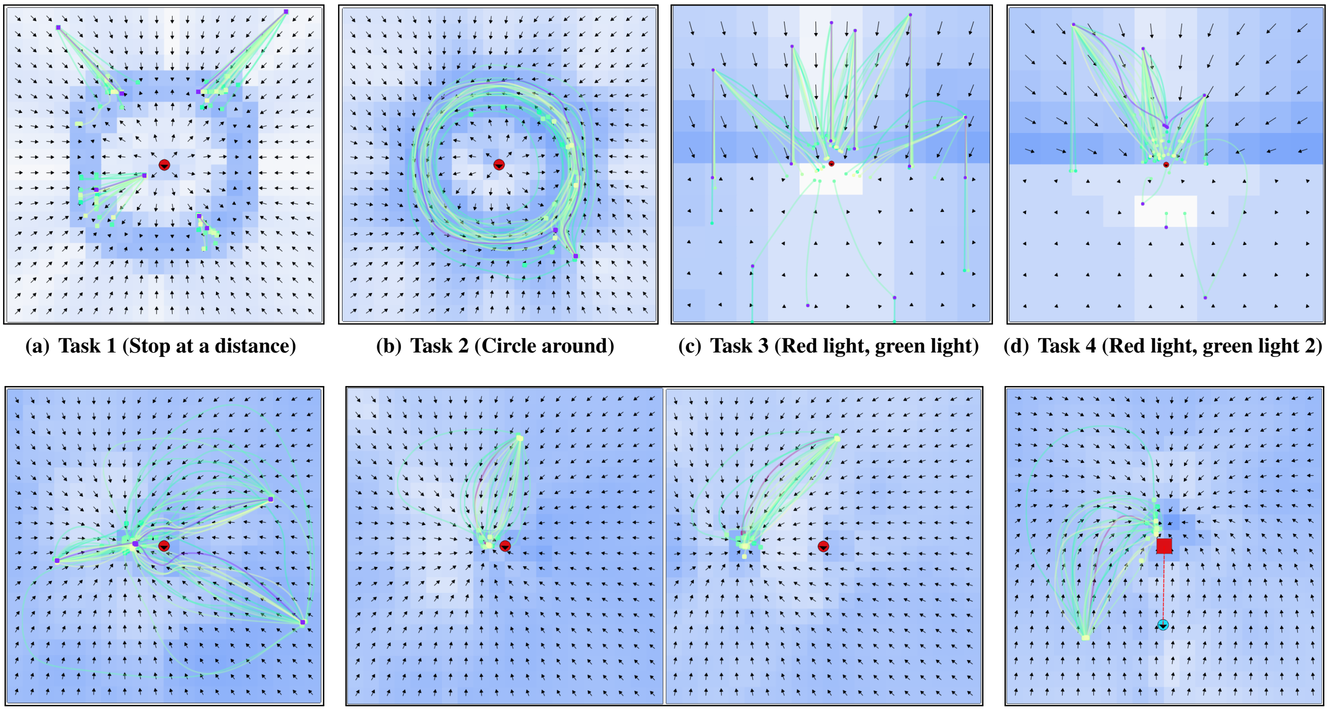
all the questions. The final usability score was then normalized to a 0-100 scale to improve readability. The overall usability score averaged over participants was 58.36±7.05 (using summation over the 1-7 scales), reaching a 80.6 percentile of the total score on the 0-100 normalized scale. When translated to the Sauro-Lewis Grading scale, the scores in this percentile receive a very high usability rating of A- (see [LS18], Table 1). As we are comparing a non-standard score of our custom questionnaire to a standardised measure of SUS, it is important to remember that the score is of a purely informative nature. Nevertheless, taken into consideration that our tool was not designed for a commercial use at this stage, and that it uses a simple interface, this grade indicates a very high usability performance.

References

[AE*16] Assila, Ahlem, Ezzedine, Houcine, et al. "Standardized usability questionnaires: Features and quality focus". *Electronic Journal of Computer Science and Information Technology: eJCIST* 6.1 (2016) 3.

[Bro96] Brooke, John. "SUS: a quick and dirty usability scale". *Usability Evaluation in Industry* 189 (1996) 2.

[LS18] Lewis, James R and Sauro, Jeff. "Item benchmarks for the system usability scale". *Journal of Usability Studies* 13.3 (2018) 4.



(a) Task 1 (Stop at a distance) (b) Task 2 (Circle around) (c) Task 3 (Red light, green light) (d) Task 4 (Red light, green light 2)

(e) Task 5 (Stay on the right) (f) Task 6 (Parametric, left = low speed, right = high speed) (g) Task 7 (Hide behind obstacle)

Figure 3: *Summary of the velocity IFs that the participants drew for all 7 tasks. The source of the fields is always the red object. The black arrow in each grid cell denotes the average IF vector for that cell among the IFs of all participants. The blue intensity of a cell indicates the variety among participants: it is the standard deviation of the Euclidean distance to the average IF vector. The green curves are the trajectories induced by the participants' fields for various starting positions. The purple curves are the trajectories induced by a 'ground truth' IF drawn by the authors before the user study. This trajectory corresponds to the video instructions given to the participants.*

Questions	Mean	SD
1. It was easy to learn to use the system.	6.36	0.66
2. I found the system unnecessarily complex. †	1.77	0.97
3. Overall, I thought the system was easy to use.	6.23	0.69
4. I found the various functions in the system were well integrated.	6.05	0.84
5. I found this system very awkward to use. †	1.68	0.89
6. I needed to learn a lot of things before I could get going with the system. †	2.36	1.18
7. It was easy to make the software do exactly what I wanted.	5.82	1.18
8. Whenever I made a mistake using the system, it was difficult to correct it. †	3.50	1.60
9. The terminology was related to the task I was doing.	5.91	1.11
10. I was wondering sometimes if I was using the right function †	2.68	1.62

Table 2: *The average ratings and standard deviations for each item of the usability questionnaire in our study. † indicate negative questions, whose score were inverted for computing the final overall score. All questions were answered on a 7-point Likert scale (from 1: Completely Disagree to 7: Completely Agree).*