We would like to kindly thank the reviewers for their constructive feedback. This document will describe the reviewers' concerns and how we have addressed them.

Reviewer BRy1

This is the second cycle of the review. I mentioned about the poor presentation of the paper in the previous round. The authors have added some comparison with another gaze model " pyStar-FC" in this version.

I am rather confused about the comparison with pyStar-FC that the authors are adding: I thought the point of the comparison is to show the proposed method produces better results in terms of human gaze or head motion - on the contrary, the authors simply mention that the proposed method can produce results similar to pyStar-FC. Then why not just use pyStar-FC from the beginning?

The authors should evaluate the motion of the head or animation that shows the improved realism of the animation due to the usage of their gaze model. There is no evaluation or qualitative test like that in the paper. The video shows some head motion of the characters watching different directions, but this is rather subtle, and it is difficult to judge if the resolution of the head area is low.

The authors mention about the potential interesting future work - these are interesting, but isn't there any evidence that this model is most suitable for such applications compared to other models?

page 6: First, it should be noted that the PSM saliency maps our models are predicated on have been previously evaluated against SALICON,

page 6: so we construction scenarios in our virtual environment

page 6: ior (inhibition of return) -> IOR

page 7: Matching it to real human gaze data should therefore be possible and is important planned future work. -> this claim sounds very optimistic and without any justification

Response: Thank you for your insights. We have added a brief explanation of why our model is preferable to using pyStar-FC alone at the end of Section 5, paragraph 3. Our model is authorable and can be tuned to match other models. The parameters provided by pyStar-FC are less intuitive to adjust, and changes to the underlying saliency algorithm would require retraining the ML model used. The comparison is to highlight the flexibility and authorability of our model. The evaluation of the head motion is a valuable suggestion. There was not time to add it, but we will take it under consideration as important future work. Whether our model is more suitable than others for any of the mentioned future work will require thorough investigation which itself is part of the future work. We have added a phrase highlighting the need for further comparison to other models to the last paragraph of Section 6. Grammatical errors have been corrected, and the sentence on page 7 has been adjusted.

Reviewer PCWz

In this paper, the authors present a gaze control framework based on a saliency map and introduce two distinct models. The first model converts a generated saliency map into a potential field, generating forces onto a simulated particle that acts as a proxy for gaze direction. The second model treats the saliency map as a probability distribution and interpolates the current gaze direction to a random point sampled from it. The authors claim that the framework is flexible and can be tuned to reproduce the performance of other frameworks.

This paper is a resubmission, and the new revision includes an additional comparison with a previous system. However, the justification provided still appears to be insufficient. The comparison with the baseline model focuses on the similarity of the fixation points computed by both models but lacks quantitative evaluation or user studies addressing the quality of the generated motion. Additionally, the discussion of the experiments is inadequate. The authors acknowledge that the models are sensitive to parameters; thus, it is crucial to provide more information about how these parameters are tuned to match the baseline. Is this tuning based on the "ten pairs of images" mentioned in the third paragraph of section 5? How does the tuned model perform on new images?

Furthermore, the models presented in the paper seem rather simplistic. The advantages of using the particle model in terms of gaze quality, as compared to a simpler approach like interpolating among saliency points, remain unclear. The work could gain significance if it demonstrates that the generated gaze motion is human-like quantitatively, or if it proposes a method for automatically tuning the parameters to match human gaze patterns. As it stands, the paper resembles a technical report and does not appear ready for publication.

Response: Thank you for your comments. We have added a discussion of the parameter tuning in Section 5, paragraph 4. Parameter tuning was done manually, which also emphasizes the authorability of the model. We also point out that automatic tuning of the model to match output of other models using an optimization framework would be fairly straightforward. We have added a brief possible justification for why our model is preferable over simpler approaches in Section 4.2.1, Paragraph 7. However, demonstrating the gaze motion is human-like and automatically tuning parameters to match would be ideal as you say, and we hope to include this in future work.

Reviewer VhB1

This paper presents techniques controlling the gaze of virtual avatars that integrate saliency maps generated by a parametric model. Two control methods are explored. The first is based on a particle method that interprets the saliency map as a potential field and a particle is integrated through the field using gradient information to gaze at minimum in the field, i.e. the most salient features. The second method interprets the intensity in the saliency map as a probability, and changes gaze targets based on a random sampling technique. There are smooth transitions between gaze targets, and a decay mechanism is used to cause target switching after some time.

I reviewed a previous version of this paper submitted for the earlier deadline. It is appreciated that the authors took the time to respond to reviewer concerns and questions in the current revised version. I am also encouraged by the authors' willingness to revise the manuscript to clearly convey the objectives and technical details of the their work.

New experiments were added that compare the behavior of the particle model to the pyStar-FC model. This not quite an "apples-to-apples" comparison since the pyStar-FC model is used to predict human gaze for real static images, whereas the proposed technique focuses on authoring gaze transitions from saliency maps. Rather I interpret this result as demonstrating how the proposed method can be integrated with existing models to generate gaze targets. However, details behind such an integration are not discussed, and it would have been valuable to know details behind how the particle model can be tuned to match the output of pyStar-FC. The manuscript should be updated to reflect these steps.

Some justification is provide at the end of Section 4.2.1 as to why the saliency map is lifted to a spline surface, since this allows interpretation as a physics type problem. I am doubtful about this reasoning. What is really needed here is an ability to find gaze targets at minima in the potential field. I suspect a robust optimization technique, e.g. particle swarms, may perform just as well or perhaps even better in cases where the potential field is multimodal or avoid issues due to coarsely sampled spline control points, e.g. the agent gazes just off-to-the side of highly salient points as can be seen in the video.

It is unfortunate that Equation 1 was not revised with a (brief) explanation of terms in the equation.

I am still on-the-fence about accepting this paper, but leaning positive.

Response: Thank you for your insights. We’ve added a clarification to Section 5, paragraph 2 that explains how pyStar-FC internally computes a saliency map before generating scanpaths. An explanation of how we manually tuned our model to match pyStar-FC was added in Section 5, paragraph 4, which we hope emphasizes the authorability of our model. We also point out that tuning the model automatically using an optimization framework would be fairly straightforward. As you suggest, robust optimization techniques such as particle swarms may produce better results, but we wanted to maintain simplicity in our model approach, however this is a great suggestion we will consider for future work. Some discussion of this was added to Section 5, paragraph 4, and to Section 6, paragraph 1. A brief explanation of the terms in equation 1 has been added to Section 4.1, paragraph 2.