Rich-cue virtual environments can be disadvantageous when discriminating navigation models

Luise Gootjes-Dreesbach[1], Lyndsey C. Pickup[1], Andrew W. Fitzgibbon[2] & Andrew Glennerster[1]

[1]School of Psychology and Clinical Language Sciences, University of Reading

[2]Microsoft Research Ltd, Cambridge

Contact: E.L.Gootjes-Dreesbach@reading.ac.uk

Introduction

Human navigation could be achieved through two distinct approaches:

- **A full 3D reconstruction of the scene.** This approach is generally used in computer vision.
- **View-based strategies**, using purely image-based features.

Here, we use a homing task in immersive VR to compare predictions from each type of model.

Homing Task

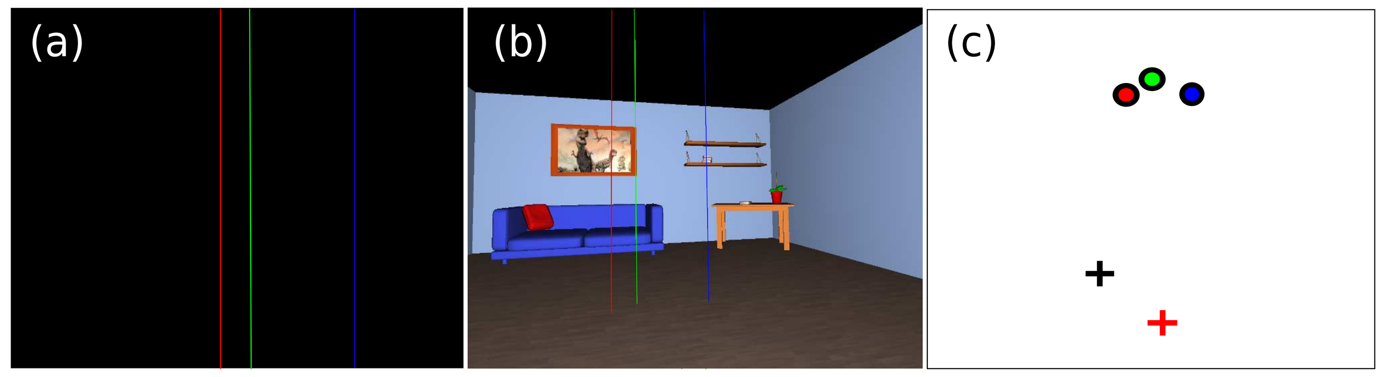

Figure 1: Example trials (a) in the sparse-cue and (b) rich-cue conditions. (c) Points are projected onto a horizontal plane for modeling. The black + marks the center of the 80x10cm viewing strip from which participants viewed the scene in Interval 1 and the red + marks the position that participants were "teleported" to in Interval 2.

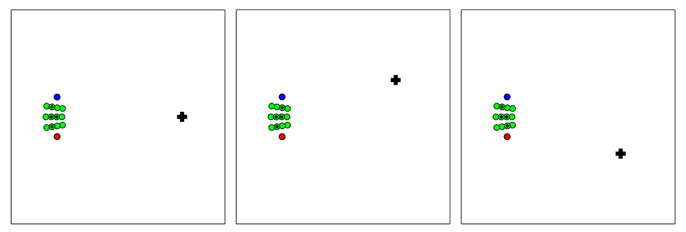

Figure 2: Middle (green) pole position and start position were varied in 48 conditions. In 36 conditions the green pole was static between intervals; in the rest it moved slightly.

Interval 1. Participants were presented with a view of 3 long thin (1-pixel) vertical poles (either in a rich cue environment or on their own). Slight lateral movement for motion parallax was allowed. A button press then set the **"goal location"**.

Interval 2. Participants were "teleported" in virtual space and had to return to the goal location. A second button press marked the **"end location"**. In some trials the green pole moved between intervals in a way that maximised the difference between model predictions.

Experimental Results

Modeling predicts the spread of errors on the horizontal plane, assigning likelihoods of the data to each point in the room (see 6). Except for S1, participants do not seem to change navigation strategy in response to green pole movement. Errors were less spread out in the rich-cue condition, showing variance mainly in distance.

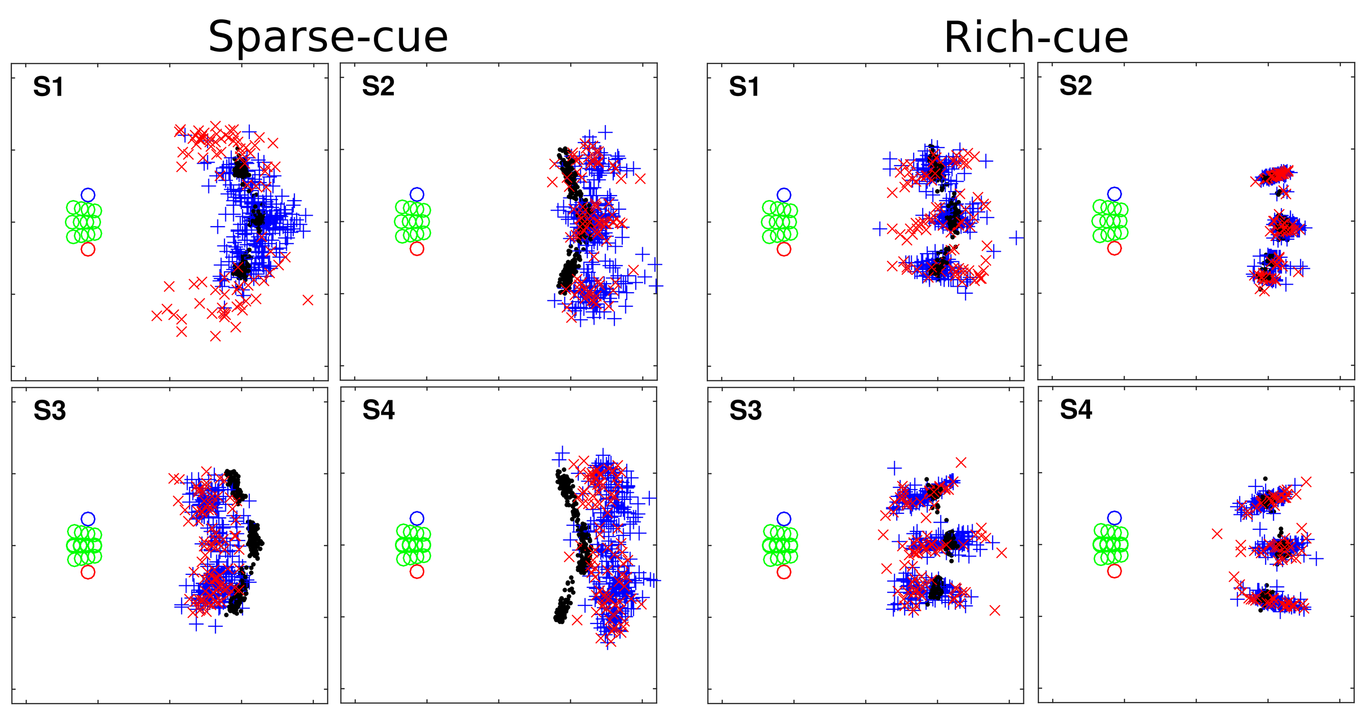

Figure 3: Spread of end points in the green pole static (blue) and green pole moving conditions (red) in the sparse- (a) and rich-cue (b) environments. Black dots show the goal points.

Modeling Methods

Reconstruction-based model

The reconstruction-based model builds a representation given a number of "cameras" along a viewing strip. The uncertainty of this representation ("image error") is represented by a 2D Gaussian per pole. To estimate how likely the current location is to be the goal location in a given trial we measure the overlap between the representations at the goal and each room location using the Bhattacharyya distance d.

The (normalized) probability of an end-point under the model is $L = \frac{1}{Z}\exp\{-\lambda d\}$. where Z is a factor normalizing the likelihoods across the ground plane. λ determines how quickly likelihoods decay based on d and is fit separately for each participant. See [2] for more details.

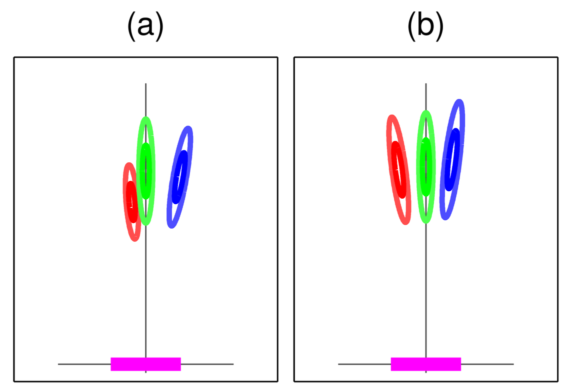

Figure 4: Representations from two different points in the virtual room (e.g. goal and end point).

View-based model

The view-based (VB) model builds representations based on two features, chosen by maximizing likelihoods in a previous data set. After fitting 2D Gaussians to summarize errors in feature space, likelihoods drawn from these Gaussians can be projected back into the virtual room. For details, see [1].

Thus, the two features, measured at the goal (G) and end point (E) are:

1. **Relative change in angle** $f_{\mathrm{A}} = \frac{\phi_\gamma^G - \phi_\gamma^E}{\phi_\gamma^G}$, where ϕ_γ is the largest angle between pairs of poles
2. **Disparity gradient** $f_{\mathrm{B}} = \frac{\delta_\alpha^G}{\phi_\alpha^G} - \frac{\delta_\alpha^E}{\phi_\alpha^E}$ where ϕ_α is the smallest angle between pairs of poles and δ_α is the disparity measured between the same two poles at the goal and end point.

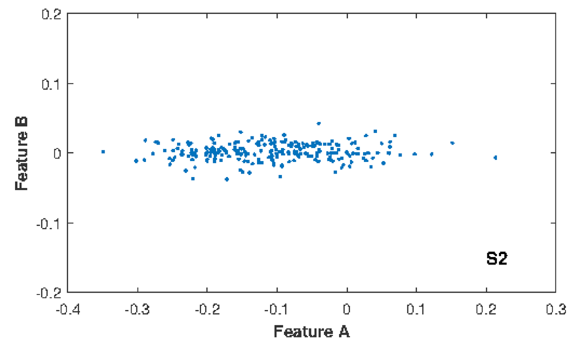

Figure 5: Feature-error space of participant S2 - each point represents the change on these features between a goal and end point. Likelihoods fit to this cloud can be projected back into the virtual room.

Model Comparison

We compared the likelihood of the experimental data against simulated data sets drawn from model predictions and compared likelihoods for all samples under both models. Training was conducted on the green pole static and tested on the green pole moving data. Data from participant S1 was trained and tested on a random split of the green pole static data, as navigation strategy may have changed in the testing set (see Figure 3).

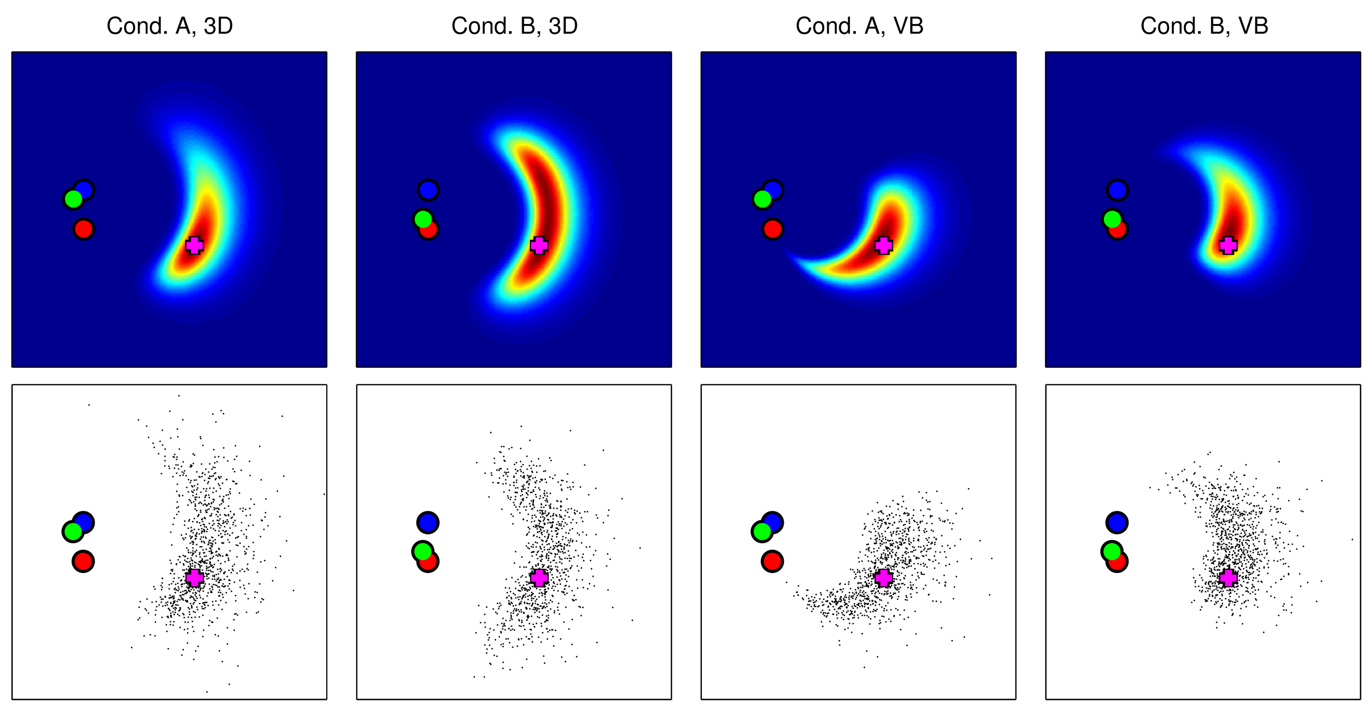

Figure 6: Sampling from model predictions: A sample from each of the 84 green pole moving trials served as a full simulated data set. 3D model on the left, view-based model on the right, each applied to two conditions.

Modeling Results

Figure 7 shows that while the reconstruction-based (3D) model fits the collected data well, it is unable to distinguish between simulated data sets drawn from itself and those drawn from the alternative model. This is not the case for the VB model. The real data lies within the view-based samples under both models.

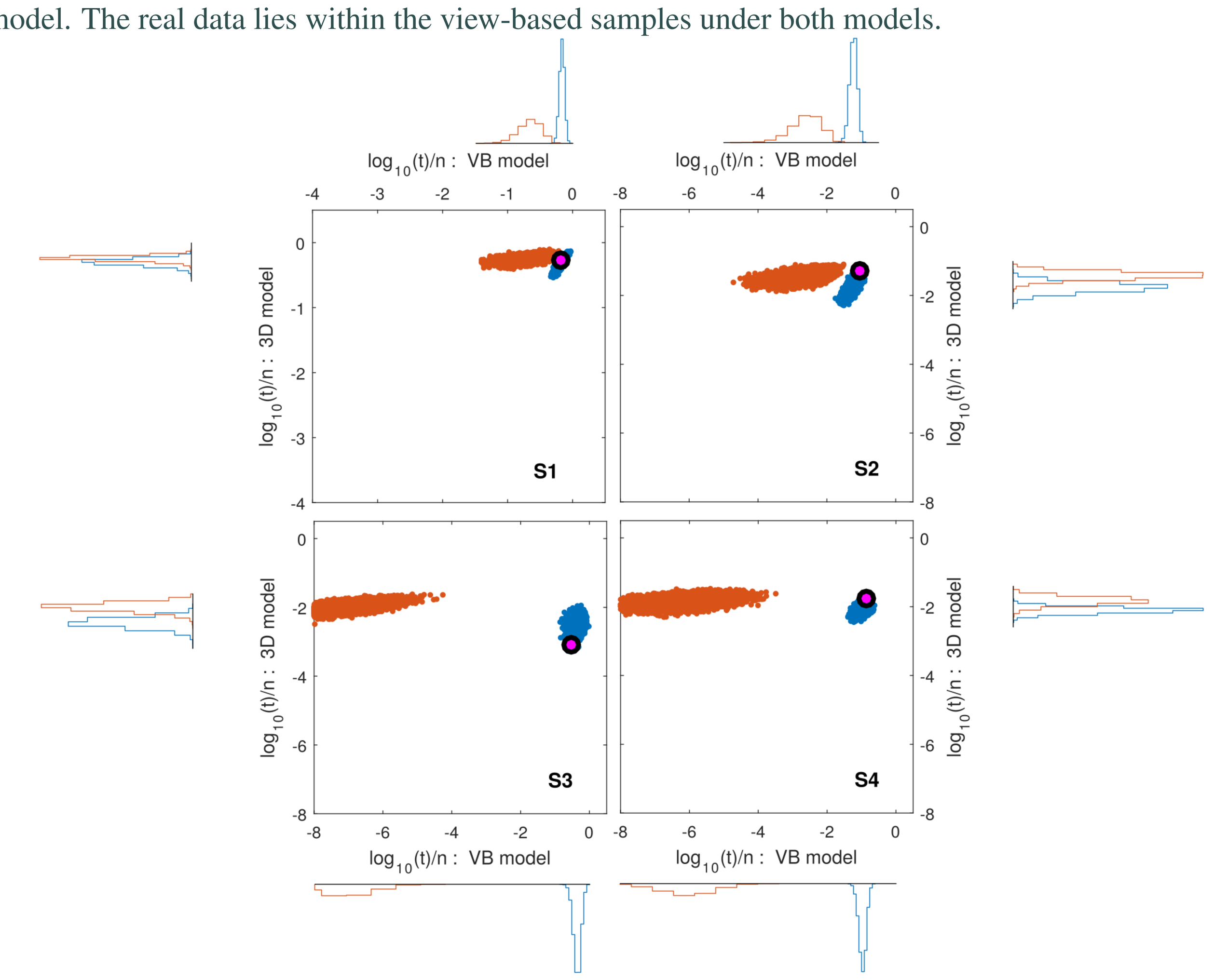

Figure 7: Likelihood of the data and of samples from the two models under each model (sparse-cue condition). Red points show the likelihood of simulated data sets drawn from the reconstruction-based (3D) model, blue points show the same for the view-based model. The magenta dot shows the likelihoods of the experimental data under each model.

Modeling of rich-cue navigation suffers from several problems: there are a very large number of possible features for the view-based modelling making a systematic search problematic. Similar issues arise for 3D modelling. Both models give rise to narrower spatial predictions than for the equivalent condition in a sparse room (see Figure 8).

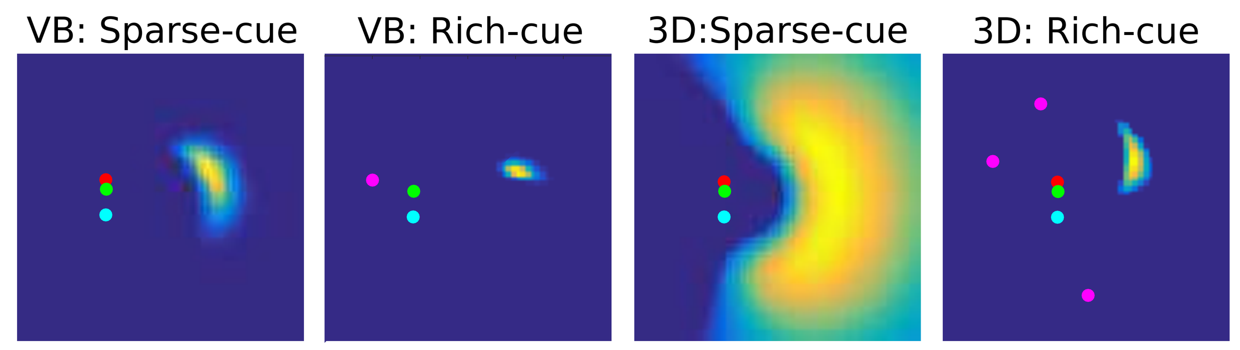

Figure 8: Predictions of the view-based and 3D reconstruction models in a sparse- and rich-cue condition.

Conclusions

- View-based model provided a better match to the data than a 3D reconstruction-based model.
- Understandably, homing is more precise in a rich-cue environment; both models predict this and it becomes harder to distinguish between models in this case.

References

[1] L. C. Pickup, A. W. Fitzgibbon, S. J. Gilson, and A. Glennerster. View-based modelling of human visual navigation errors. *IVMSP Workshop*, pages 135–140, June 2011.

[2] L. C. Pickup, A. W. Fitzgibbon, and A. Glennerster. Modelling human visual navigation using multi-view scene reconstruction. *Biological cybernetics*, 107(4):449–464, 2013.