Computational Cognitive Science
9.660 Final Project
December 19th, 2011
Professor Joshua Brett Tenenbaum

Nikola Otasevic
(otasevic@mit.edu)

Tracking object under occlusion

Introduction

One of the challenges in computer vision is part segmentation and action recognition. Many approaches solve the problem in constrained conditions (under certain object properties, illumination, scales etc.), but all of these seem far from what people seem to do when they recognize parts and actions. People have no problem with scales, partial occlusion in space or time, blurry images or different illumination.

In the process of understanding how human vision deals with these problems, I decided to pursue a project whose purpose is to reveal how people track objects in the presence of occlusion and how well they are able to generalize patterns of motion that they observe during the part of the motion in which an object is not occluded.

I designed an experiment in which I varied different parameters that are likely to affect how people track objects. Experiment consisted of tracking a ball in one-dimensional motion. Subjects would always observe the ball in motion before it would go into a tunnel. During its motion in the tunnel, subjects were asked to determine the most likely position of the ball.

This is the story that was given to the subjects prior to the experiment:

> ### *Where is the ball?*
>
> *Your task is to estimate a position of the ball every time you hear a short, high pitch sound.*
>
> *You will be given situations in which a ball is going to run into a tunnel and you will be asked to estimate a position of the ball at a particular time during its motion in the tunnel. You will always see the ball before it goes into the tunnel. When you hear a short,*

high pitch sound, the ball stopped. Then you should use your mouse and click on the most likely position of the ball. You can take as much time as possible, but you can click only once for every situation. When you mark your estimate of the position of the ball with a mouse click, you will be presented with a new setup. When you are ready, click on the button "Start", track the ball and wait for the high pitch sound again.

There are a total of 33 situations. Feel free to take a break in between any two situations. Total experiment time should be about 10 minutes.

After reading the story, subjects were presented with the following setup of the experiment.

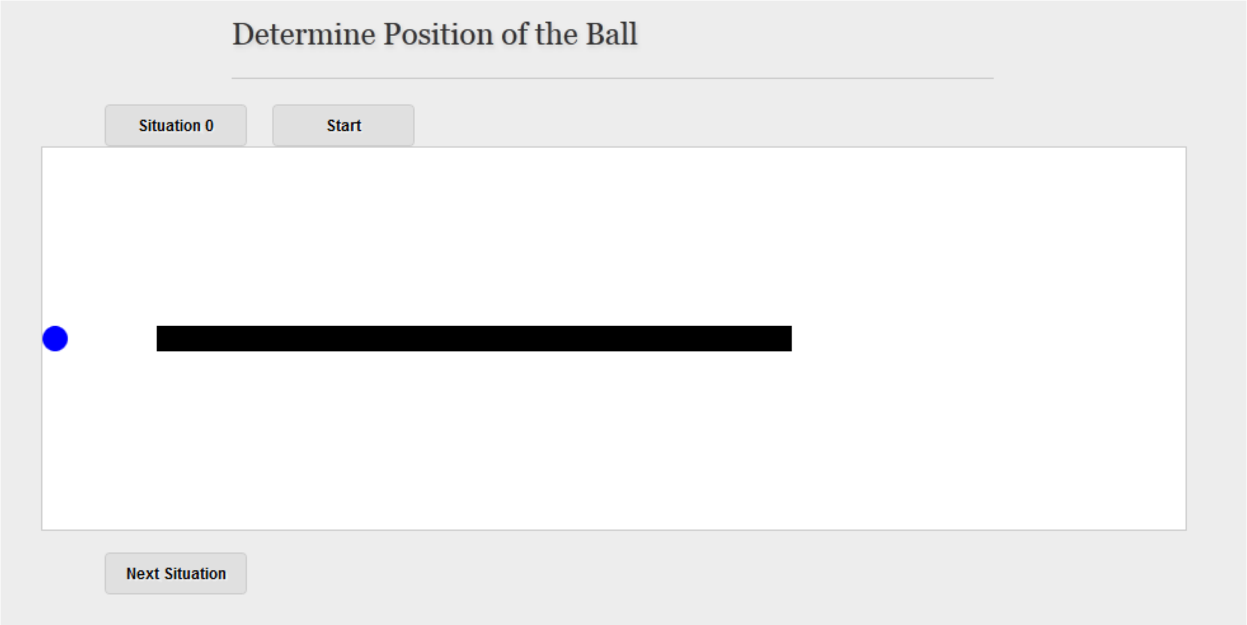


Figure 1: Experimental setup

In an attempt to understand how different parameters of the motion affect the estimates, I varied the velocity, acceleration and position of the occluder. In addition to that, I also varied the position within the occluder when the ball stopped. This produced a total of 32 different situations. Subjects were given one additional situation (situation 0) which served only as a way of introducing the tasks they would need to perform throughout the experiment.

Model of single object tracking with occlusion

The simplest and most intuitive model that we can use to think about this situation is presented in the following figure:

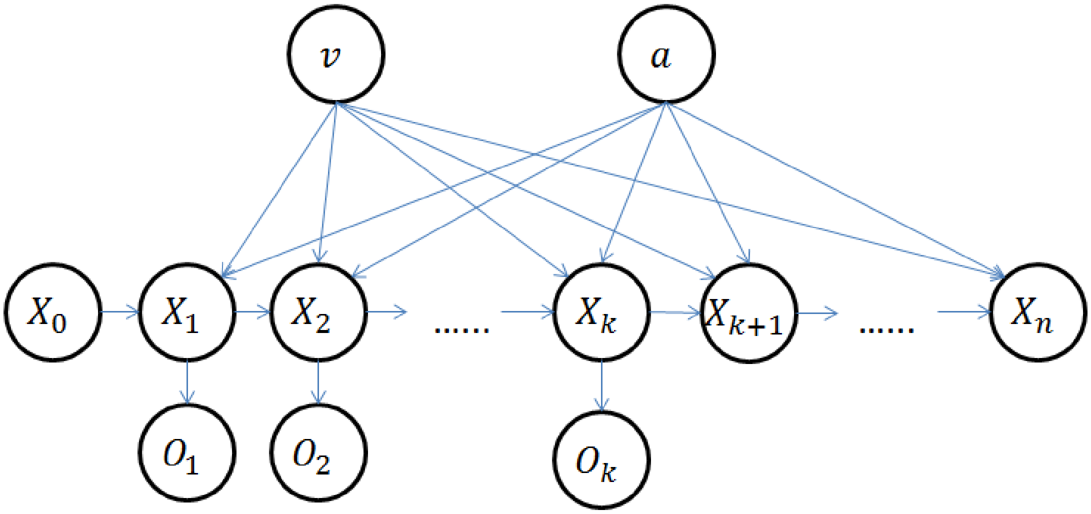


Figure 2: Model of the situation in which people assume that initial velocity and acceleration are constant throughout motion. First k positions are observed, while positions from $k+1$ to n are occluded.

In this model, positions X_1 through X_k are observed by the subjects, while initial velocity and acceleration are hidden parameters. Based on the observed motion, subjects are asked to estimate positions during the occluded part of the motion. The dynamics of the model are given by:

$$X_s = X_{s-1} + v * \delta t + a * t * \delta t + \frac{a * (\delta t)^2}{2}$$

where δt is the time between two positions and t is the time between positions X_0 and X_{s-1}.

Any three observations of states are enough to estimate both acceleration and initial velocity:

$$a = \frac{O_s - 2 * O_{s-1} + O_{s-2}}{(\delta t)^2}$$

$$v = \frac{O_s}{t} - \frac{t}{2} * \frac{O_s - 2 * O_{s-1} + O_{s-2}}{(\delta t)^2}$$

Since observations are noisy, we can expect that these calculations will vary. I decided to model the acceleration and velocity with a Normal distribution in which mean is adjusted with the new observations:

$$P(A_t) = N(\mu_{At}, \sigma_A)$$

$$\mu_{At} = \sum_{i=3}^{t} \frac{(O_i - 2 * O_{i-1} + O_{i-2})}{(\delta t)^2}$$

Similarly, for velocity we have:

$$P(V_t) = N(\mu_{Vt}, \sigma_V)$$

$$\mu_{Vt} = \sum_{i=3}^{t} \frac{O_i}{i} - \frac{i}{2} * \frac{(O_i - 2 * O_{i-1} + O_{i-2})}{(\delta t)^2}$$

This kind of dynamics assumes that longer observation of the ball yields more accurate estimates of the parameters of the motion. These estimated values for A_t and V_t are then used in predicting positions of the ball during the occluded part of the motion.

The last step in modeling this problem is to model the observations of positions of the ball that humans make. I modeled this with normally distributed noisy parameter:

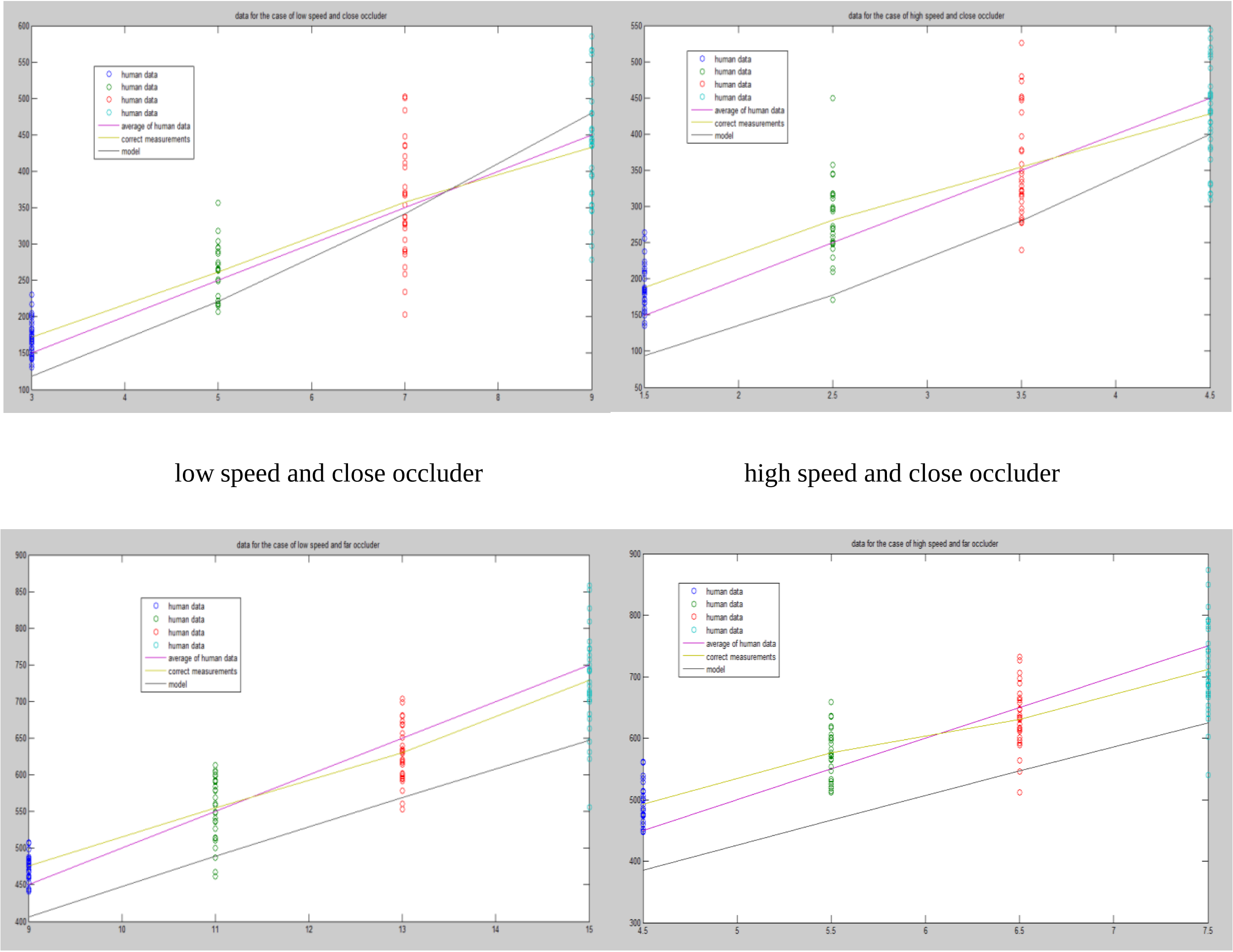
$$O_t = X_t + N(0, \sigma_o)$$

Data Analysis and Comparison to Model

Variation in velocity

In this experiment, I tested 30 subjects and obtained 30 samples, one per each subject.

On the following figure we see how variation in velocity of the object and position of the occluder affect human and model's ability to accurately estimate position of an occluded object.



low speed and close occluder

high speed and close occluder

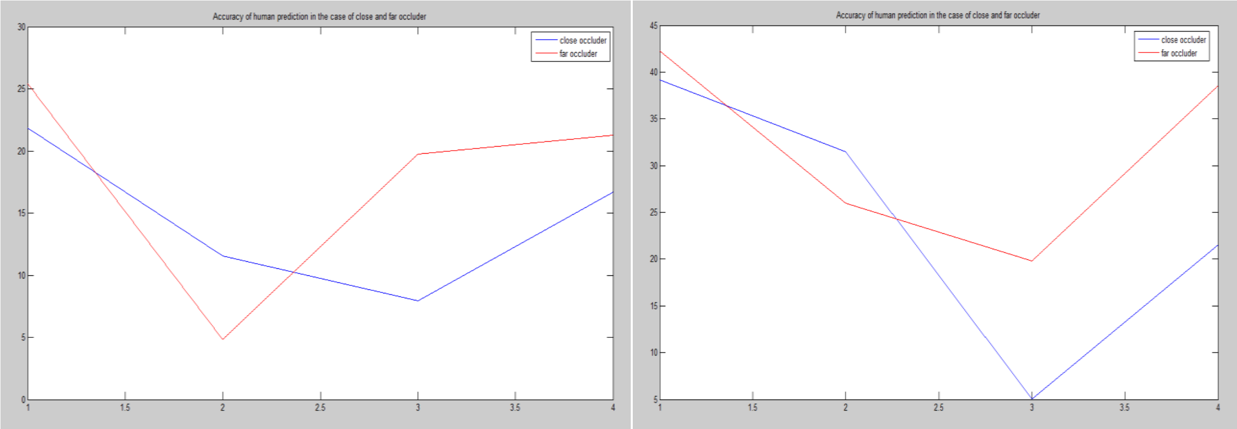
low speed and far occluder

high speed and far occluder

Figure 3: Plot of human predictions (yellow) against true predictions (purple) and model's predictions (green)

As we can see, higher velocity tends to distort ability to correctly predict position, while position of the occluder does not seem to be significant in this case. Model typically predicts slightly lower positions. That could be explained by the fact that velocity and acceleration in the model are estimated jointly which causes model to believe that there is some small acceleration which typically produces slightly lower velocity.

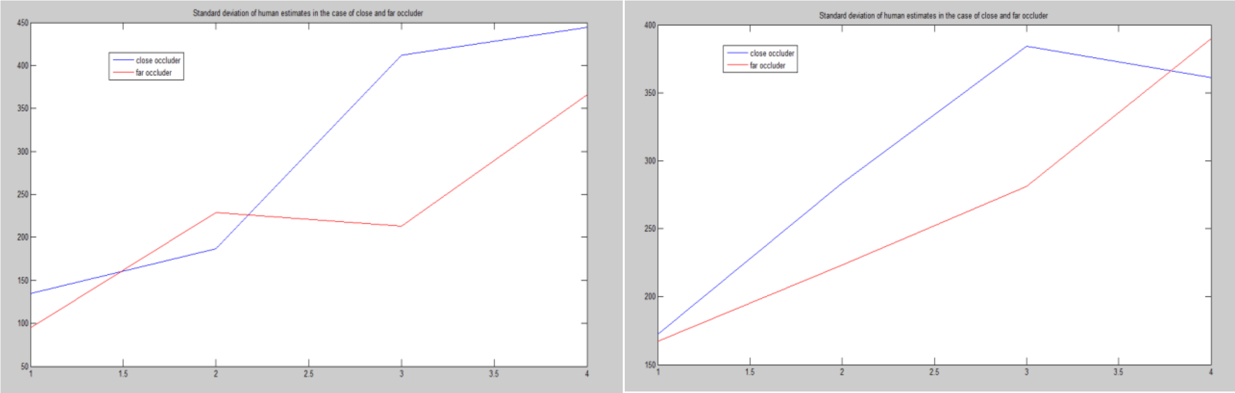
In the following four figures, accuracy and precision of human estimates are compared in a more analytical way using absolute error as a measure of accuracy and standard deviation as the measure of precision.



Low speed

High speed

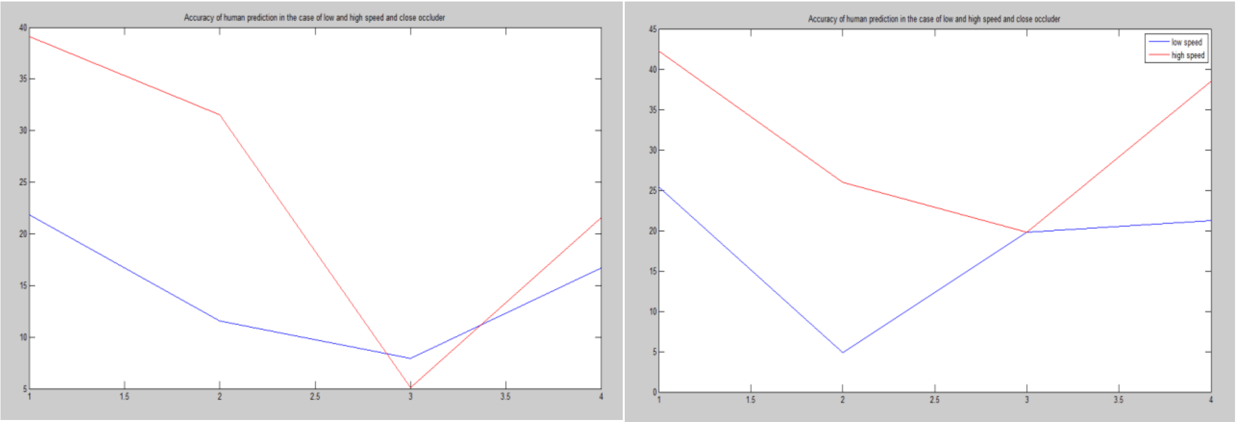
Figure 4: Comparison of absolute error of human prediction – variation in the position of the occluder. Red and blue lines represent far and close occluder respectively.



Low speed

High speed

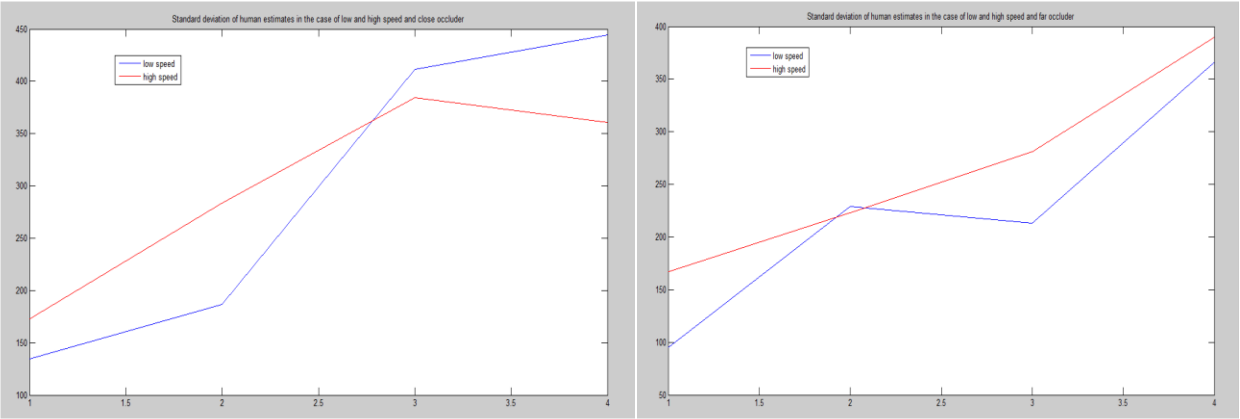
Figure 5: Comparison of std. deviation of prediction – variation in the position of the occluder. Red and blue lines represent far and close occluder respectively.



Close occluder

Far occluder

Figure 6: Comparison of absolute error of human prediction – variation in the speed. Red and blue lines represent high and low speed respectively.



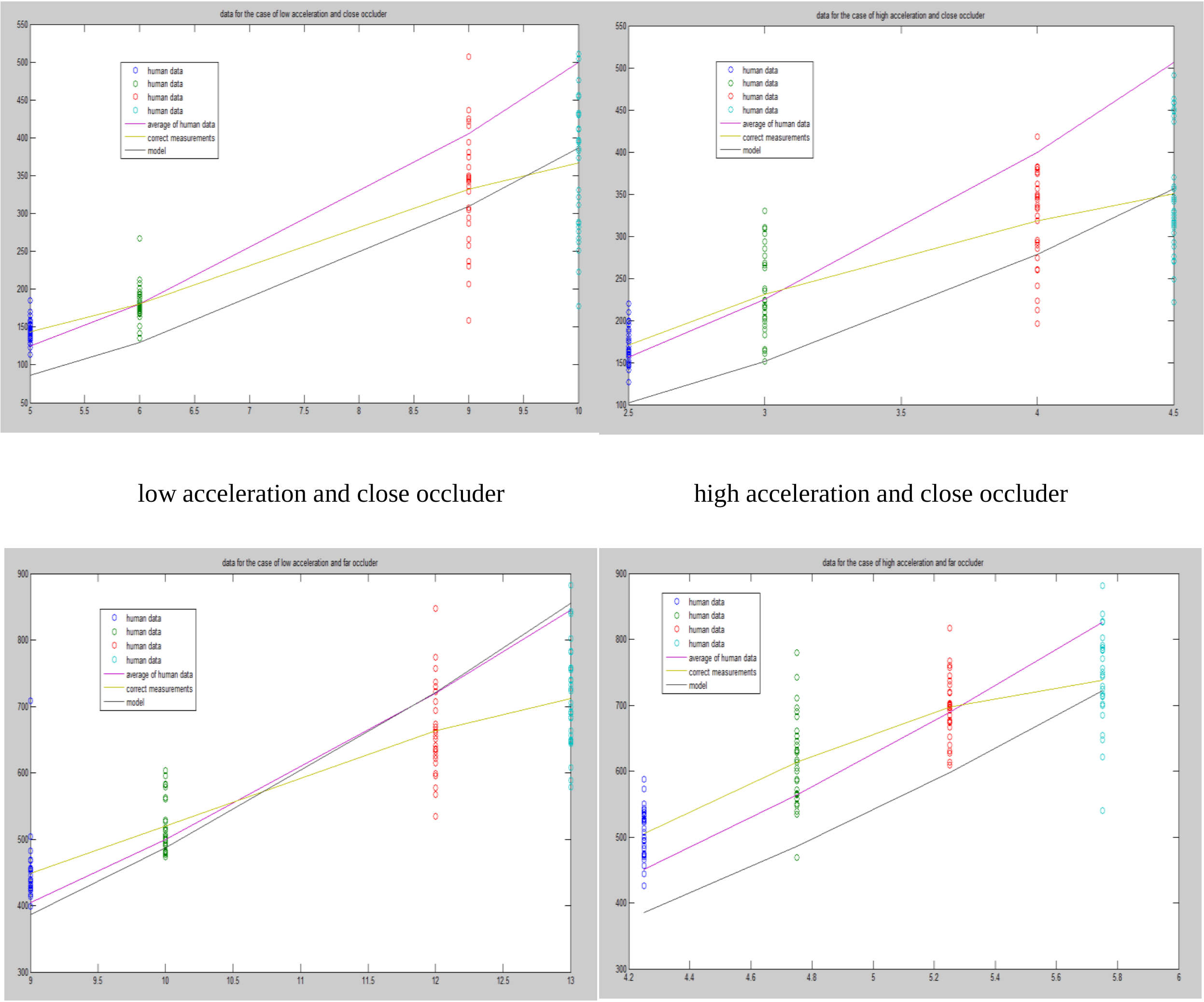
Close occluder

Far occluder

Figure 7: Comparison of std. deviation of human prediction – variation in the speed. Red and blue lines represent high and low speed respectively.

Variation in acceleration

On the following figure we see how variation in acceleration of the object and position of the occluder affect human and model's ability to accurately estimate position of an occluded object.



low acceleration and close occluder

high acceleration and close occluder

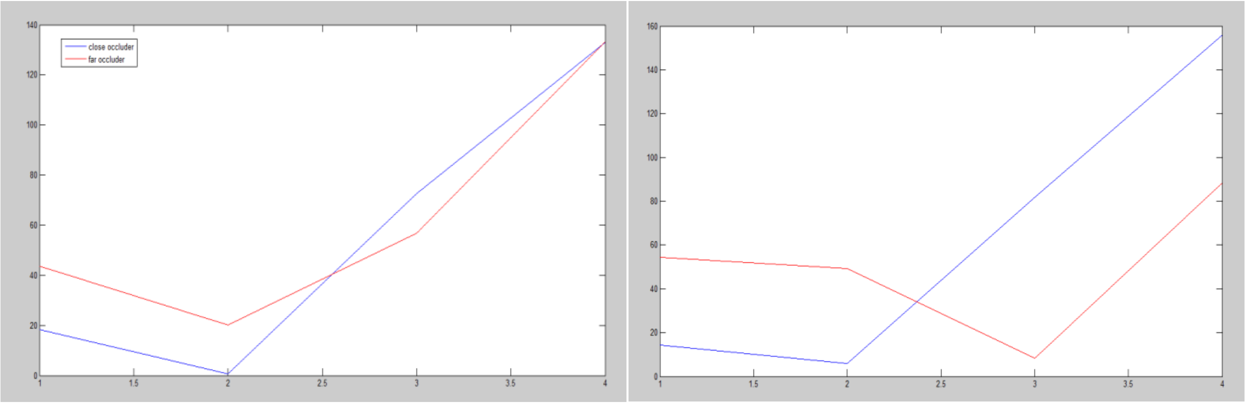
low acceleration and far occluder

high acceleration and far occluder

Figure 8: Plot of human predictions (yellow) against true predictions (purple) and model's predictions (green)

We can see that model behaves well compared to the correct states, but not so well compared to human predictions. Interesting trend in human data is apparent deceleration across all 8 conditions which cannot be explained by the offered model. In the analysis, I will offer a few hypotheses and specify how this model could be extended to account for pattern of deceleration in human observations.

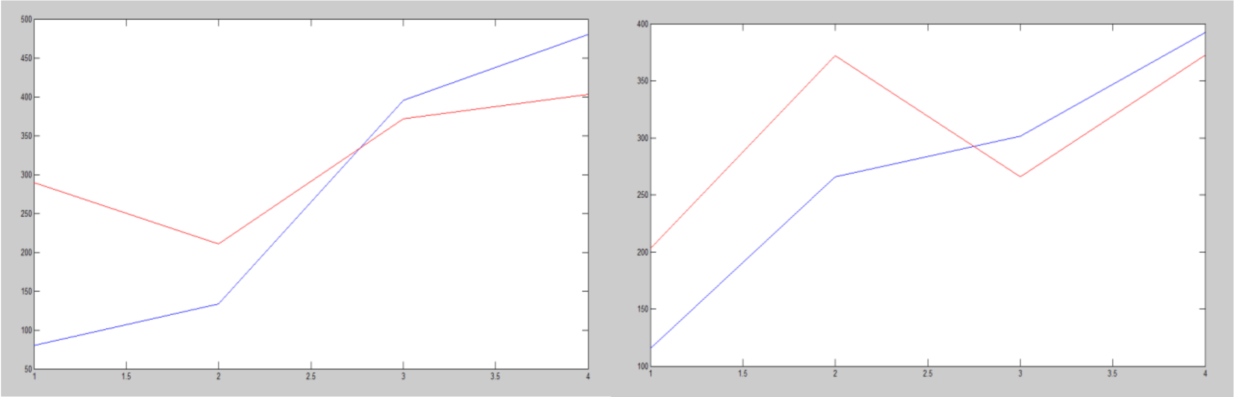
In this experiment, variation in acceleration of the object as well as variation in the position of the occluder did not show any statistically significant impact on the accuracy (absolute error) and precision (standard deviation) of human estimates. I demonstrate this more formally on the following 4 figures.



Low acceleration

High Acceleration

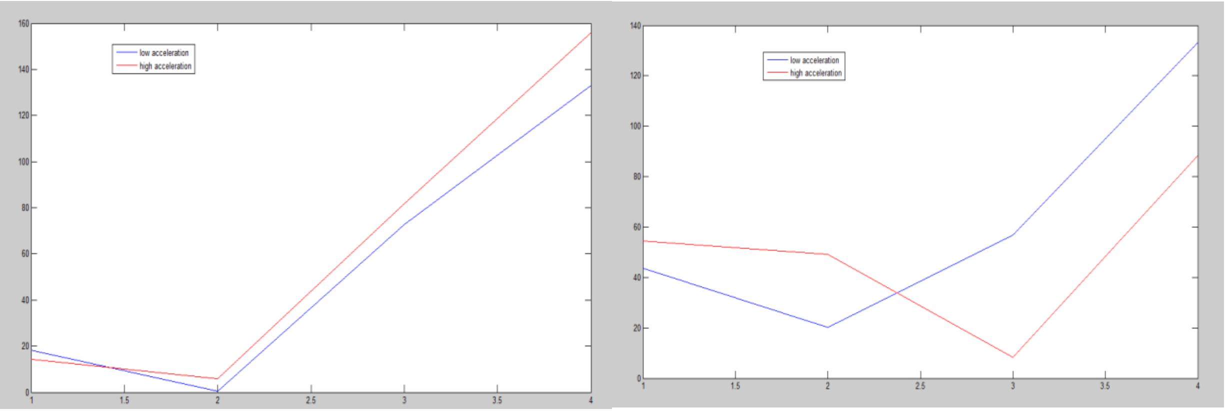
Figure 9: Comparison of absolute of human prediction – variation in the position of the occluder. Red and blue lines represent far and close occluder respectively.



Low acceleration

High Acceleration

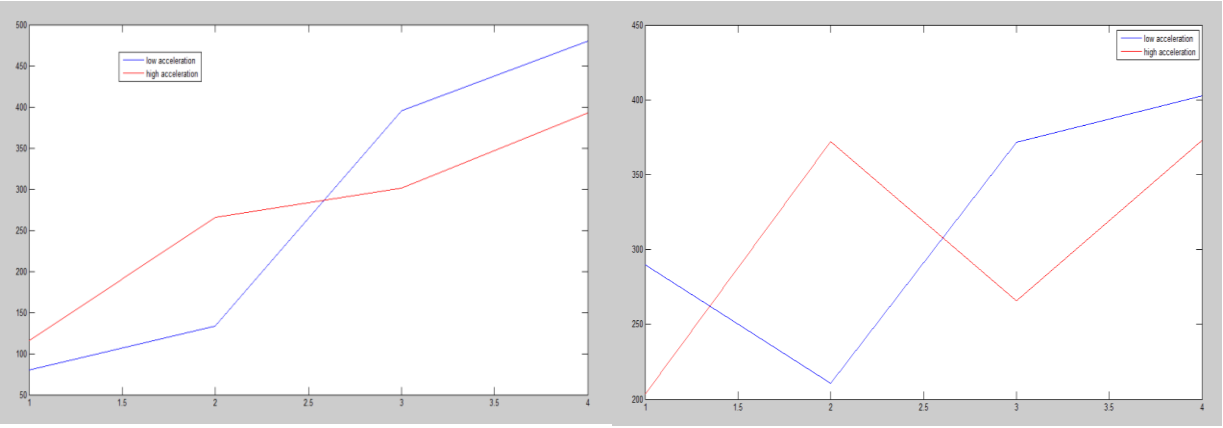
Figure 10: Comparison of std. deviation of prediction – variation in the position of the occluder. Red and blue lines represent far and close occluder respectively.



Close occluder

Far occluder

Figure 11: Comparison of absolute error of human prediction – variation in the acceleration. Red and blue lines represent high and low acceleration respectively.



Close occluder

Far occluder

Figure 12: Comparison of std. deviation of human prediction – variation in the acceleration. Red and blue lines represent high and low acceleration respectively.

Interesting trends

First very interesting trend that we can see in the data is the fact that people tend to systematically overestimate position of the object very soon after it is occluded. When we look at human estimates of the first position (the one closest to the left end of the occluder), we see that vast majority of data points (almost all) are above the correct position, while the variance is relatively small. That means that people are actually very certain about their predictions, but they are wrong.

Another interesting trend that we can recognize in the data is that slope of the curve seems to be decreasing suggesting some kind of decelerating pattern in the position of the ball. This is very unintuitive considering that subjects' expressed beliefs about pattern of the motion were usually correct. That is, when asked, they typically expressed correctly that they are seeing some kind of acceleration or constant motion, but never expressed a belief that there might be some decelerating pattern. However, data reveals different pattern.

Hypotheses

There are a few hypotheses that could explain these two unexpected trends:

- **"Fear" of predicting the position close to the boundary**. This hypothesis attempts to explain both phenomena at once by simply having a model in which there is a very low prior probability of assigning position to the values that are close to the boundaries of the occluder.

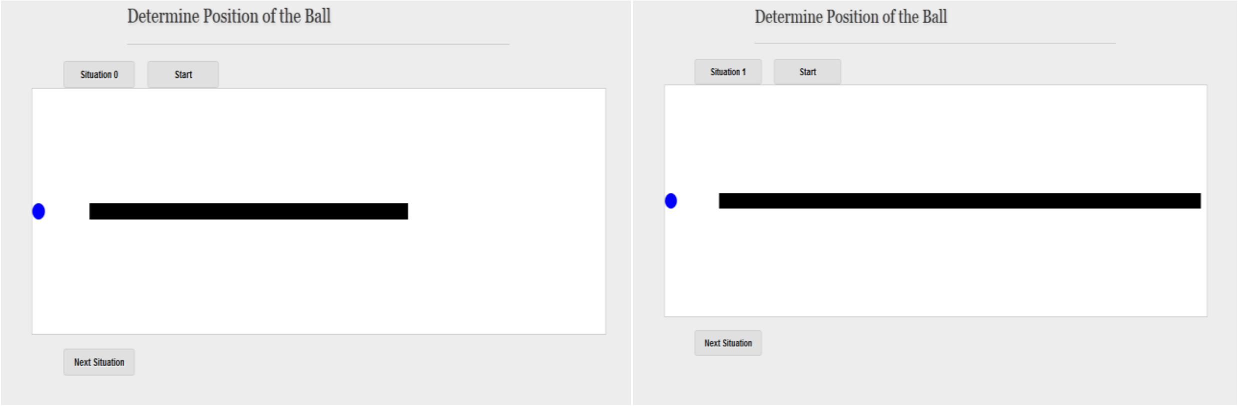
  On a cognitive level, this could happen because of the existence of two mechanisms of tracking – mechanism for the real tracking (while seeing the object) and mechanism for imagining the object while it is being occluded. These two mechanisms are fundamentally different. They are very likely to engage completely different amount of cognitive power and compute the tasks in a very different manner. It could be that switch between these two mechanisms requires certain time which is why humans unconsciously assign very low probabilities to positions right after or right before the switch.

- **Different time scale**. This hypothesis attempts to explain the decelerating pattern of motion that people unconsciously predict. The assumption is that time scale exhibits a log-like pattern in which the equivalent intervals of time get represented as smaller and smaller in human brain. This would imply that with the unchanged estimates of the initial velocity and acceleration, estimates of the changes in the position decrease with the real world time. However, this hypothesis alone does not explain the systematic overestimate of the positions in the beginning of occlusion.

  On a cognitive level, different time scale might be the consequence of the delay which could be a consequence of overload. As I mentioned before, imagining the motion of an object under occlusion is likely to be a resource-demanding operation which might introduce large biases in other tasks such as time keeping. Another explanation is that this kind of scale is simply hard wired in humans and that its correction was not evolutionary necessary.
- **Animated objects do not have self-propelling mechanisms**. This hypothesis indicates that even though they expressed belief that objects are not decelerating, people unconsciously believe that animated objects such as ball on the screen do not have ability to accelerate without an external force being applied to them. That is, they have certain initial velocity and acceleration, but if we let them go, they will stop after certain time due to some kind of friction.

Additional Data

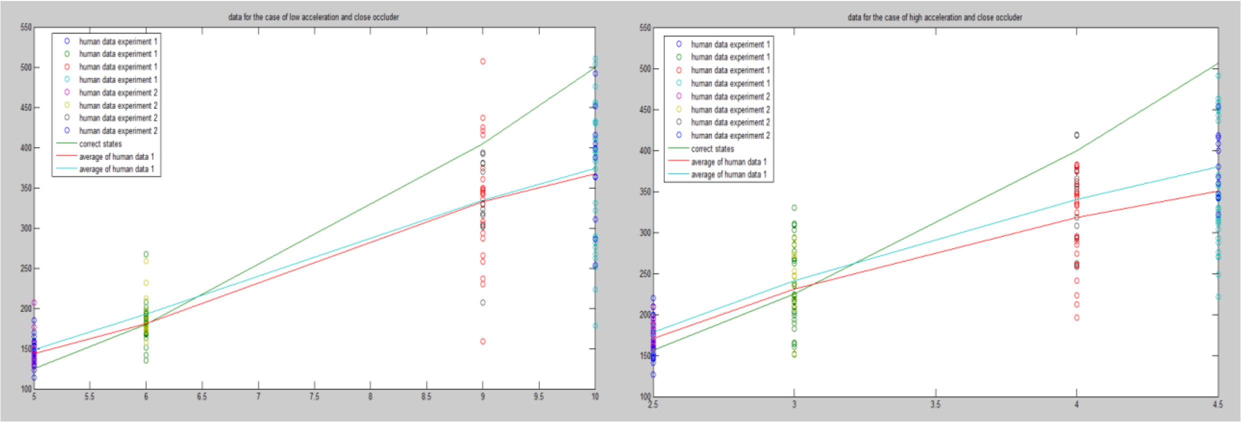
In order to test the first hypothesis ("Fear" of predicting the position close to the boundary), I collected more data. This time, in the situations that involved close occluder, the length of the occluder was 800 pixels instead of original 500. The points at which the stop sound was played remained the same.


Determine Position of the Ball
Situation 0
Start
Next Situation
Determine Position of the Ball
Situation 1
Start
Next Situation

Original situation in the experiment Modified situation in the experiment

Figure 13: New Experiment Setup

Prediction of the "Fear" hypothesis is that human estimates at the positions near the end of the original occluder are not going to be underestimated.



low acceleration and close occluder high acceleration and close occluder

Figure 14: Plot of human predictions in experiment 1 (red) and human predictions in experiment 2 (blue) against true predictions (green) in the case of close occluder and variation in acceleration

However, as we can see on the two plots presented above, this hypothesis does not seem to generate significantly different results, but it seems to offer a partial explanation. As we can see, blue curve (experiment 2) is constantly above the red curve (experiment1) which suggests that people have some tendency to bias their estimates based on the duration of occlusion.

Hypothesis about animated objects can be modeled with additional parameter (γ) added to the model specified before:

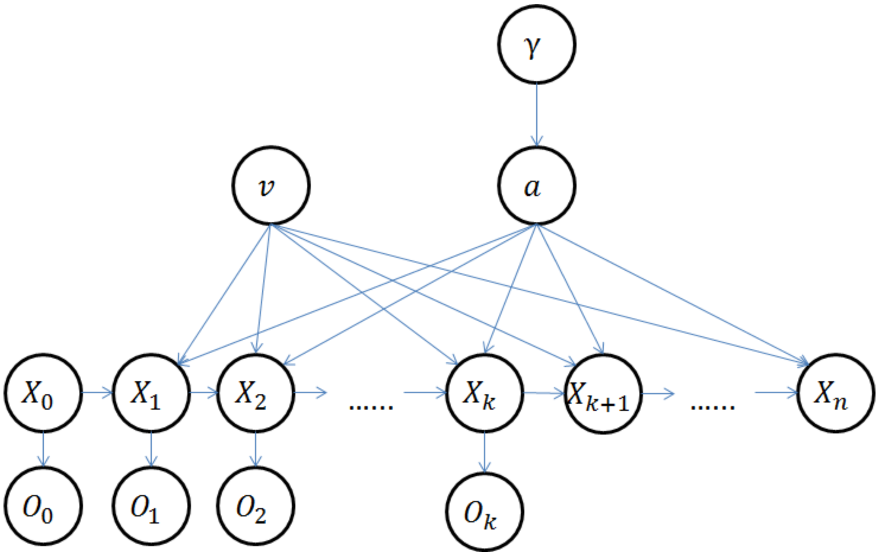


Figure 15: Augmented model with animated object parameter γ

Parameter γ in the augmented model specifies the likelihood that a given object has some kind of self-propelling mechanism that can be used to accelerate. This can also be thought of as a prior distribution on acceleration parameter given the look of the object and situation. This hypothesis can also be easily tested by changing the look of the situation. Firstly, we can observe the acceleration in the y-axis direction in order to see whether people would associate that kind of acceleration with the gravitational force. Other way of doing this would be to make an object look like a car starting from the zero position at the traffic light. Both of these would increase γ and its effects could be measured. Another quick way to test this hypothesis would be to provide a surface which looks like it might have a higher friction coefficient and test whether human estimates suggest even higher deceleration. I did not test any of these in this project.

Finally, hypothesis about the log-like time scale would certainly yield better results on this particular experiment; however, it would be much more interesting to see whether this hypothesis extends to longer periods of time under occlusion which again I did not have time to test for in this project.

Conclusion

In the attempt to get closer to modeling how human vision apparatus works, I concentrated on the part of the human vision that deals with objects under occlusion.

I designed and implemented an experiment to test how accuracy and precision of object tracking under occlusion depend on different parameters of motion such as velocity, acceleration and time spent in motion before the occlusion. For the details of implementation please refer to the attached zip file. For the demo of the experiment, go to http://web.mit.edu/otasevic/www/demo.html.

I implemented a model of object tracking under occlusion and compared it to the results that I obtained from the experiments. In that comparison, I discovered two interesting trends in the data obtained from the experiments:

- tendency to overestimate positions of the object soon after it was occluded
- tendency to estimate decelerating pattern in motion in spite of expressed belief about accelerating pattern

Finally, I offered three hypotheses that could explain the existence of these two patterns. I tested one of those and gave a brief guide for how the other two might be tested.