FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Surya Prakash Chembrolu

Predicting accuracy in Multiple Object Tracking tasks from trajectory statistics

Department of Software and Computer Science Education

Supervisor of the master thesis: Mgr. Filip Děchtěrenko, Ph.D.
Study programme: Computer Science
Study branch: Artificial Intelligence

Prague 2023

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I will be so grateful to my supervisor M.Sc. et Mgr. Filip Děchtěrenko, Ph.D. for his kind and friendly approach in dealing things and excellent guidance he provided throughout this project. I would like to thank all my teachers who taught me till now. Special mention to my study advisors Solange Petracchi and Kristyna kysilkova for their encouragement and support over all the years. Lastly, I am so blessed to have my parents who provided me benefits and showered so much love as always.

Title: Predicting accuracy in Multiple Object Tracking tasks from trajectory statistics

Author: Surya Prakash Chembrolu

Department: Department of Software and Computer Science Education

Supervisor: Mgr. Filip Děchtěrenko, Ph.D., Department of Software and Computer Science Education

Abstract: Cognitive science is an interdisciplinary area covering neuroscience, psychology, linguistics, philosophy, and computer science. Computer science and cognitive science mutually benefit from each other because computer science is very helpful to design and perform experiments in order to understand how the brain works likewise research output from cognitive science can lead to new concepts and models in artificial intelligence. Within cognitive science, Multiple Object Tracking (MOT) paradigm is used to study visual attention. In MOT experiments, participants are required to keep track of some moving objects in parallel. In this study, a data-driven approach is taken in order to explain the tracking performance of the subjects taking part in MOT experiments. The stimuli in MOT known as trajectories or tracks presented in previous studies were taken and the difficulty of those trajectories is quantified based on trajectory statistics. Then a model is created to explain tracking performance and this model is tested in an MOT experiment.

Keywords: Multiple Object Tracking, Prediction, Modelling,Visual Attention

Contents

Introduction

Visual attention is an important cognitive process that helps us to focus on relevant visual information while suppressing the noise in the environment. Imagine different situations like driving a car or playing some team sport, visual attention plays a crucial role in decision-making whether it is stopping the car if a pedestrian is crossing the road or making a pass to another player in the team who is in a better position to score a point. When involving eye movements we can differentiate attention as *covert* meaning directing attention to a specific entity in the scene without moving the eyes onto it and *overt* meaning attentively moving the eyes to focus on a particular object in the scene. Attention can even be differentiated as *selective* if there is a single focus of attention like for example, while reading some text, and *divided* is if there are multiple foci of attention like for instance listening to the news on the radio while driving. Divided attention helps us to multitask by processing multiple stimuli simultaneously. In order to understand how the brain processes multiple stimuli and the limitations of those processes attention has been widely studied in Neuroscience and Cognitive psychology. Advancements in technologies like fMRI, EEG, Eye tracking devices, and of course computers helped researchers to come up with various methods to study attention in more detail.

Pylyshyn and Storm(1988) introduced an experiment called Multiple Object Tracking(MOT) to study divided attention where participants are asked to keep track of a subset of objects that are cued for example by color or flashing those objects before the experimental trial begins and when the trial begins these objects usually move in a randomized way somewhat resembling Brownian motion. Once the trial ends participants need to select the objects they tracked. Different studies are conducted to understand factors impacting tracking performance i.e, how many objects could the participants track accurately in the trials they are presented. Intriligator and Cavanagh(2001) showed tracking capacity is affected by spatial resolution. Tracking performance decreased when objects moved in smaller frames compared to objects that moved in larger frames on the screen. In a dynamic task like MOT, tracking capacity is impacted by the speed of objects moving as shown in (Alvarez and Franconeri, 2007) where observers could track only one object when objects with extreme speeds are presented and they could track up to eight objects at very slow speeds. Alvarez and Cavanagh(2005) came up with the Hemi-field effect showing that tracking performance decreases when objects to be tracked are in the same hemifield and increases when tracking objects are in different hemifields. The accuracy is reduced when tracking time in the experiment is increased((Oksama and Hyönä, 2004)).

MOT being a dynamic task was useful to researchers who tried to understand the relationship between eye movements and attention. Zelinsky and Neider(2008) came up with two gaze strategies that participants use which are target looking and centroid looking where participants adapted different strategies based on the number of objects they are asked to track. Fehd and Seiffert(2008) performed MOT experiment with three to five targets and reported that participants preferred to use centroid looking more during tracking compared to target looking strategy. Later a study (Fehd and Seiffert, 2010)tested target looking strategy with the combination of centroid/target-looking strategy and showed that there was a significant drop in tracking performance when only target looking strategy was used. Lukavsky (2013) though primarily examined the inter and intra-individual differences in eye movements of the subjects also studied the connection of eye movements to the strategies mentioned above. Findings from this study showed that much of the eye movement variability is because of the anti-crowding strategy where the observer or participant tries to reduce crowding which is a visual phenomenon that occurs when it is difficult to separate target stimuli from distractions close to it. Dechterenko and Lukavsky(2016) used neural networks to predict gaze positions and compared them to gaze predictions from the analytical strategies mentioned. In this study, they found that neural network predictions were better than the predictions from strategies.

In our study, we wanted to take a data-driven approach to predict tracking performance just based on the trajectories presented(stimulus in MOT task is usually referred to as trajectory or track). We used data collected from a previous experiment (Děchtěrenko et al., 2017) to quantify the difficulty of trajectories that were presented based on trajectory descriptors, and then we created a model with two metrics that would explain tracking performance in MOT task. We conducted an experiment presenting a classical MOT task with 8 objects and tested the model on the data collected.

In Chapter 1 we discussed some background about the MOT task, factors influencing tracking performance, and MOT models that explain tracking limitations. In chapter 2 we discussed some aspects of predictive modelling, we introduce the data used in this work, findings from our analysis done prior to formulating the model, and how the model is created.

In chapter 3 we describe the experimental design and procedures carried out to collect data in order to validate the model.

In chapter 4 results from the experiment we conducted are discussed. Followed by a discussion in chapter 5 and the conclusion.

1. Vision and MOT

In this chapter we briefly discuss some aspects of visual processing, provide some background about MOT task, discuss limitations that affect performance during tracking and go into few frameworks that explain the limitations during multiple object tracking.

1.1 Visual processing

The eyes, visual cortex, and pathways between the eyes to the visual cortex play a vital role in vision. Light reflected by things around us enters the eye through the pupil and the amount of light that enters through the pupil is regulated by the iris which controls the size of the pupil. In the eye, there is a lens that focuses the object or scene onto the retina, and the image formed on the retina is inverted due to the lens. Within the retina, we have two types of photoreceptor cells known as rods and cones. Low levels of light and motion are processed by rods which are located on the periphery of the retina. Cones are located in the central region of the retina and they process details like color.

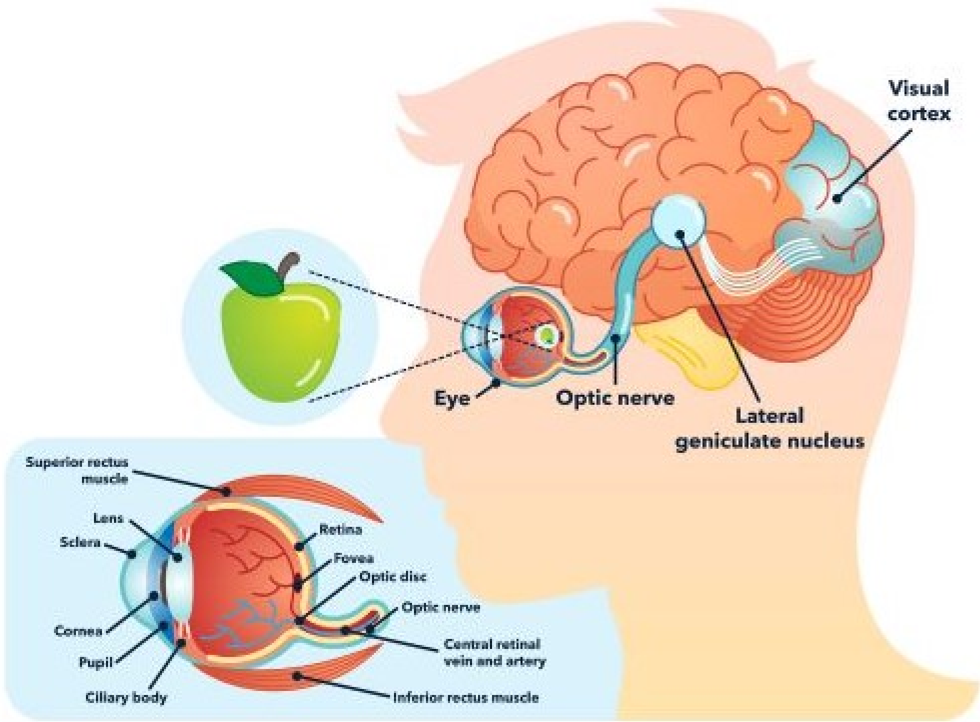

Figure 1.1: Human vision system [Source:https://www.perkins.org/wp-content/uploads/1970/01/the_brain_14.png]

The optic nerve sends visual information to the primary visual cortex via optic chiasm which is like an intersection point for information entering from both the visual fields and where this information is split into parts for further processing. From the optic chiasm, the sensory information enters the Lateral geniculate nucleus(LGN) and then send to the visual cortex for higher-level processing. In humans and also animals sensory information passing from the retina to the visual cortex is processed at a different stage in a multi-layered and hierarchical structures which actually inspired the creation of early convolution neural networks(CNN's) which are widely used in artificial intelligence.

1.2 Visual angle

When we are looking at two objects of different sizes there is a chance they can have the same size on the retina because the size of the image projected on the retina is determined by the real size of the object and the distance between the object and the eyes. To avoid this ambiguity, a measure called visual angle is used to compute the ratio between object size and distance from the eye. Visual angle is formulated as follows:

$$\theta = 2\arctan(\frac{s}{2d}) \tag{1.1}$$

Where s is the size of stimuli, d is the distance from the center of the stimuli to the eye and θ is stimuli size in degrees of visual angle.

1.3 Multiple object tracking

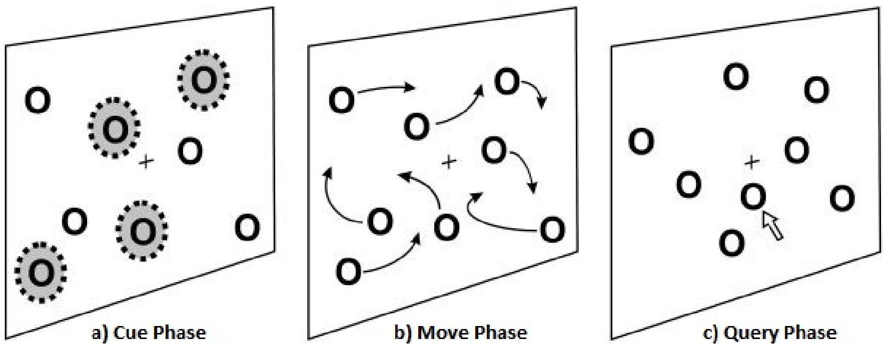

Figure 1.2: Multiple object tracking task

In each trial of MOT task n objects are presented and consist of three phases as shown in Figure 1.1, the first phase is called the *cue* phase(1.1a) where a subset of objects x_i(where $i = 1..m < n$) namely *targets* are highlighted or distinguished from remaining objects $x_j(m \leq j < n)$ which are referred to as *distractors* by flashing those objects or using some color to mark them. The second phase is known as *move* phase(1.1b) where usually all objects look identical and start moving in a randomized way on the screen. Once the objects stop moving on the screen then comes the third phase called *query* phase(1.1c) where participants are asked to select the objects that were highlighted in the *cue* phase which is nothing but the set of target objects m. Traditionally MOT is presented with 8 objects out of which 4 objects are *targets* and this configuration is denoted as 4:4 and generally, it is denoted as m:(n-m).

1.4 Neuroscience of MOT

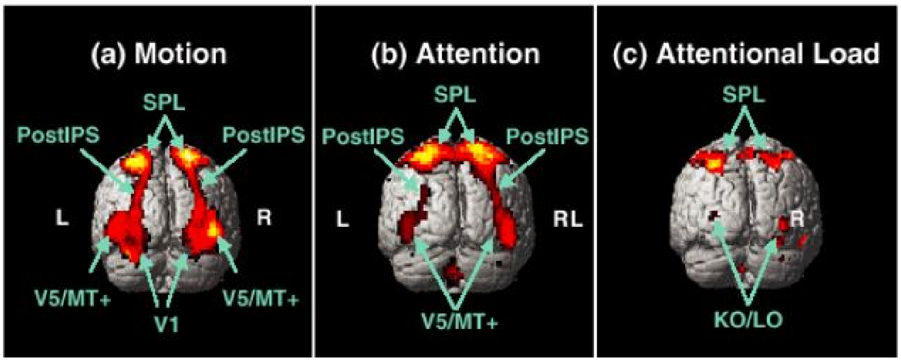

Figure 1.3: fMRI images showing activated brain areas during MOT task[Source: (Jovicich et al., 2001)]

Before heading to psychological theories explaining MOT first let us discuss some brain areas that are involved during participation in MOT tasks. From fMRI studies conducted(Culham et al.,Jovicich et al.) it is understood that broad dorsal network including areas in the posterior parietal cortex and in the frontal cortex are regularly activated during tasks involving motion.In another study where participants tracked moving targets, it was revealed that the brain area called the posterior intraparietal sulcus(PostIPS in figure 1.2a, 1.2b) associated with visual working memory is involved in the tracking task. Further research indicated increased activity in the MT+ area(figure 1.2a, 1.2b), parietal areas(posterior and anterior intraparietal sulcus and superior partial lobule(SPL in figure 1.2a,1.2b), and frontal eye fields in the frontal cortex during tracking tasks.

Some of the brain regions mentioned above responded in a dissociable manner when MOT task components are manipulated (Xu and Chun, 2006; Shim et al., 2010). For example, activity in the posterior parietal cortex correlated with an increase or decrease in the number of targets whereas activity in frontal eye fields correlated with rotational speed and the number of targets to be tracked (Shim et al., 2010). Based on research done in order to understand attention it is suggested that the role of the posterior parietal cortex may be in allocating and maintaining the spatial indices and frontal eye fields may play a role in suppressing eye movements or in maintaining higher levels of attention.
In cognitive science, it is important to understand the structures and functions of different brain areas in order to create a model or framework explaining some psychological processes that occur in the brain.

1.5 Performance limits

In this section we will look at different factors that affect the tracking performances of the participants, the percentage of trials in which participants selects all target objects correctly is defined as accuracy. Earlier studies that are done to get insights about tracking performance have identified some factors like capacity, speed, crowding, and hemifields that impact tracking performance.

Capacity

In this context, it is important to mention that visual working memory plays a critical part in MOT tasks. Capacity can be defined as how many objects can a participant keep track of during the experiment. George Miller(1956) claimed that working memory capacity is limited to 7 ± 2 items. Early studies using MOT have suggested that on average participants could track around four or five objects at once. Taking evidence from brain studies, Cowan(2001) showed that in short-term memory capacity to store 4 objects is very consistent.

Speed

Speed is evidently a limitation in any dynamic task, this limitation seems to be clear when looking at some of the physical constraints brain has. Alvarez and Franconeri (2007) showed that sometimes participants could track only a single object and at times when object speeds are set to extreme low values participants tracked upto eight objects.

Crowding

Crowding can be defined as the adverse impact of nearby entities when focusing on region of interest (Levi, 2008). Crowding is a visual phenomenon that impairs perception. Studies on crowding identified some factors that determine if crowding occurs and main factor that determines crowding is ratio of spacing between targets, distractors and the gaze point (Bouma, 1970).

Mathematically, Crowding can be defined as follows,

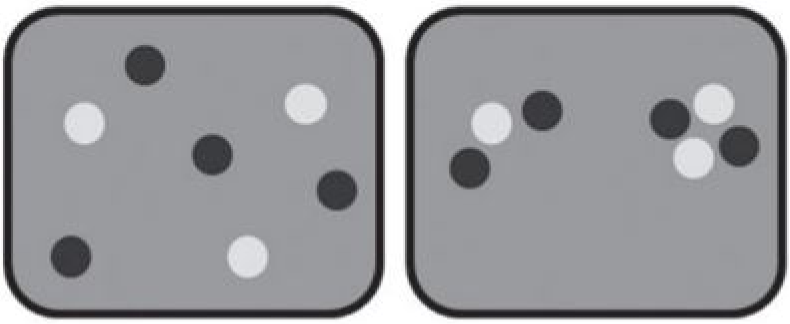

Figure 1.4: Crowding effect[Source: (Scimeca and Franconeri, 2015)]

$$\sum_{t \in T} \sum_{d \in D} \frac{||f - t||}{||t - d||} \tag{1.2}$$

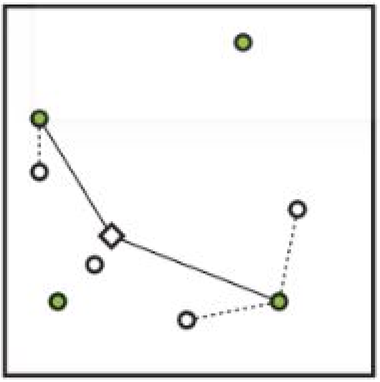

Figure 1.5: [Source: (Lukavskỳ, 2013)]

Where T represents a set of targets(green dots in figure1.4) and D represent a set of distractors(white in figure1.4) In equation 1.1 for each target object, we are calculating the ratio of the distance between the target and the gaze point(diamond point in figure1.4) to the distance between the target and the distractor. From MOT studies with eye tracking(Lukavskỳ, 2013) showed that participants use an anti-crowding strategy where they try to minimize the effect of crowding by changing their gaze point and this can be represented as follows,

$$\min_{f \in F} \sum_{t \in T} \sum_{d \in D} \frac{||f - t||}{||t - d||} \tag{1.3}$$

All models varying from low-level anatomical models to high-level attentional models have proposed that crowding occurs during the feature integration phase in the visual field if features are not combined properly when one field is loaded with more features.

Hemifields

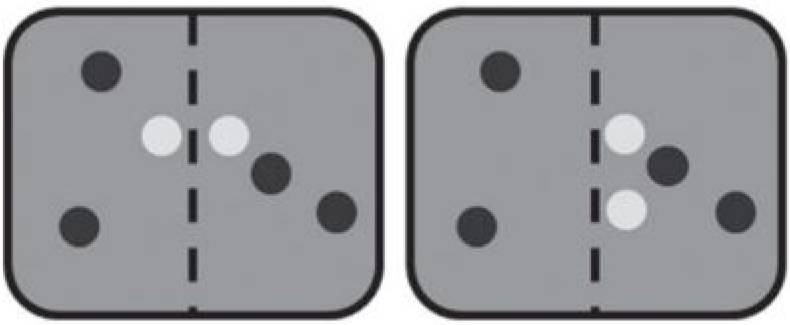

Figure 1.6: Hemifields effect[Source: (Scimeca and Franconeri, 2015)]

Anatomically brain is separated into left and right hemispheres and information coming from the left half of the visual field is processed in the visual cortex of the right hemisphere and vice versa. Alvarez and Cavanagh (2005) showed that there are exceptions to integrated processing of the two hemifields and argued that tracking performance is limited by object positions relative to visual hemifield boundary(dotted line in figure1.3). The performance seemed to be better when target objects are distributed among the hemifields and performance dropped when they are all present in the same hemifield. Performance limits discussed above lead to an underlying mechanism known as limited processing resources. In further sections below, three classes of limited processing resources namely Spatial, Temporal, and Shape are discussed.

1.5.1 Spatial Limits

Pylyshyn and Storm(1988) along with introducing MOT paradigm also mentioned a tracking mechanism that uses pointers(this can be thought of as some stick that is pointing to something) which helps the visual system to identify, attend and track the set of target objects. This mechanism accounts for a high-level representation known as an attentional priority map(Serences and Yantis, 2007). A priority map is a cortical representation that marks a set of selected locations in the visual field for enhanced processing. Spatial interference theory Franconeri et al. (2010) suggests that participants find it hard to discriminate targets from other objects when distractors fall within the inhibitory area surrounding the targets. In this studyFranconeri et al. (2008), they showed that the speed of objects might not constrain tracking resources but tracking performance is solely dependent on the inter-object spacing. Crowding and hemifield effects mentioned earlier can also be explained in terms of spatial limits.

1.5.2 Temporal Limits

Temporal limits are mainly addressed in terms of attentional spotlights and serial operation during tracking. In the absence of parallel processing, there is an attentional spotlight that is used to shift the focus of each target object and update its location. With an increased number of targets, the attentional spotlight needs to be shifted between more objects and this will affect tracking performance. When object speeds are increased the process of updating target locations could be uncertain or wrong and lead to a decline in performance. Increasing both the number of targets and object speeds compounds in impairing performance.

1.5.3 Shapes

This class of processing resources assumes there is a shape recognition system that supports tracking in a given task. During tracking this system treats a set of moving objects as vertices of an imaginary polygon. Yantis(1992) showed that tracking multiple objects might involve complex representations like a polygon that is formed by the target's position at a given instance. Though this could be a plausible argument, unfortunately, there is not much scientific evidence to support this argument. Shape formation and maintaining the shape can be operated more efficiently when the target objects are spread across the hemifields compared to these objects when they are in the same hemifield could explain the hemifield effect. The ability to form complex shapes depending on the capacity demands of the task can affect the tracking performance. Shape recognition can also suffer based on the object's speed as it affects shape maintenance over time.

1.6 MOT Models

This section covers different theoretical frameworks that have been proposed to explain the limitations in MOT tasks. These frameworks are presented in the chronological order of research. We discuss the drawbacks of some frameworks and how are these addressed in other frameworks that came in later.

1.6.1 Visual Index Theory(FINST)

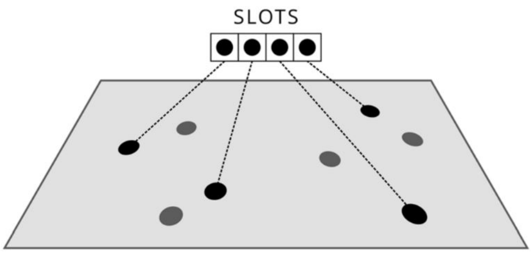

Figure 1.7: FINST theory[Source:(Meyerhoff et al., 2017)]

Visual index theory(figure 1.3) also known as FINST theory was the first theoretical framework proposed to explain MOT, this theory assumes that early vision has a visual index mechanism that connects objects in the real world to the mental image of those same objects and this connection is established by pre-conceptual visual indexes called FINSTs(FINgers of INSTantiation) which provides references to the objects(1989,2001). Visual indices maintain the connection with the indexed object during motion or eye movements whether the objects look identical or not allowing tracking of multiple objects in parallel.

1.6.2 Perceptual Group Model

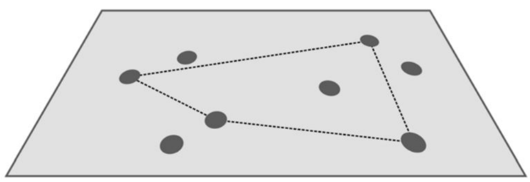

Figure 1.8: [Source:(Meyerhoff et al., 2017)]

The grouping model(figure 1.4) was the second framework explaining MOT, this theory (Yantis, 1992) proposes that the visual system creates a higher-order visual representation of the individual target objects shown in the stimulus like forming an imaginary polygon linking the target objects. The evidence for the statement of this theory is that attention is space-based and not object-based as in the previous theory, in space-based attention at a given point there is only a single convex spatial locus of focussed attention((Hoffman and Nelson, 1981)), and all entities inside the locus are attended. Yantis(1992) in his work showed that tracking accuracy was more when the higher order object like polygon remained in shape and tracking accuracy is less when the shape collapsed.

1.6.3 Multifocal Attention Theory

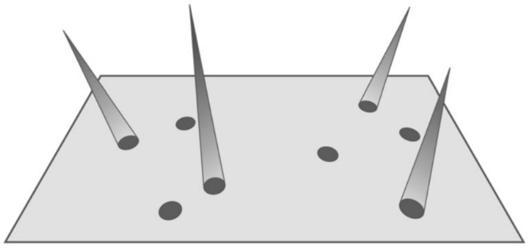

Figure 1.9: [Source:(Meyerhoff et al., 2017)]

In previous two theories discussed above considered only a single focus of visual attention or selected attention. In order to explain MOT in terms of divided or distributed attention Cavanagh and Alvarez(2005) proposed multifocal attention theory(figure 1.5). This model suggests that during the trial there are multiple foci of attention that keep track of the objects of interest. Multiple foci of attention act in a similar way to visual indices in FINST theory. However, the visual indices in FINST theory can access only one object at a given point but in this theory, there is always continuous access to all the objects being tracked. Neuroscientific(Awh and Pashler, 2000) and behavioral(Müller et al., 2003) research showed that the attentional focus can be split and the processing of stimuli at independent locations is enhanced.

1.6.4 FLEX Model

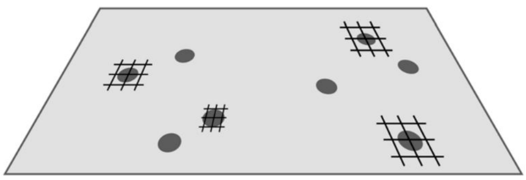

Figure 1.10: [Source:(Meyerhoff et al., 2017)]

In all three models discussed above tracking ability is limited by the number of visual indices or attentional foci which can be referred to as architectural constraints. Models with fixed architecture can predict only a set of objects that do not exceed the architectural constraints. Alvarez and Franconeri in 2007 investigated whether there is a fixed limit on the number of objects that can be tracked or whether the tracking limit can be changed flexibly depending on the type of MOT task. Considering the capacity and space tradeoffs in their work they showed that participants could track up to eight objects when speed is significantly slow and could only one object when speeds are very high. Based on these observations they introduced the FLEX model(figure 1.6) in which attentional resources are flexibly allocated between the objects during tracking(demand-based allocation). According to this model, tracking performance is limited when there is a deficit in attentional resources to attend all the targets.

1.6.5 Spatial Interference Theory

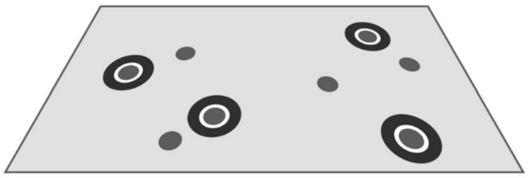

Figure 1.11: [Source:(Meyerhoff et al., 2017)]

According to spatial interference theory(figure 1.7), attentional enhancement is created and an inhibitory zone is formed around the objects being tracked. As explained in (Franconeri et al., 2010) tracking performance degrades when distractors interfere with the inhibitory zones of targets or when the inhibitory surround of one target interferes with the attentional enhancement of another target. Experiment results have shown that performance limiting factors like the number of objects to track, number of distractors, or object speeds can all be traced back to the spatial changes happening between the moving objects. This theory was successful in explaining the variances in tracking performance and has sparked a lot of research in the direction of inter-object spacing to understand tracking performance.

2. Modelling

In this chapter we briefly discuss some principles of statistical modelling, data that we worked with is described, discuss some findings from our initial data exploration and close with how the model is formulated.

2.1 Statistical Modelling

Modelling is a mathematical approach to analyzing and understanding the relationship between response variables and explanatory variables in the dataset. Statistical techniques like regression can be used to identify these relations and make predictions about future outcomes. Statistical modelling is used in diverse areas like economics, psychology, biology, and engineering.
General principles of statistical modelling are discussed in the sections below are presented with reference to Dobson and Barnett.

2.1.1 Exploratory data analysis

EDA is performed as the first step in order to get insights about the data. Usually, different data visualizations like histograms, scatterplots, heatmaps, etc are used during this process based on the type of analysis that interests the investigator. EDA helps us to identify patterns or understand relationships between different variables in the data before making any assumptions to create a model.

2.1.2 Model Formulation

Model has two components:
1. Probability distribution of response variable.
For example, $Y \sim N(\mu, \sigma^2)$
2. An equation linking the expected value of the response variable with a linear combination of the explanatory variables.
For generalized linear model the equation is as follows,

$$g[E(Y)] = \beta_0 + \beta_1 x_1 + ... + \beta_m x_m$$

where right side part of the equation is called the *linear component.*

2.1.3 Parameter estimation

For classical or frequentist statistics the most commonly used estimation methods are least squares and maximum likelihood. Both of these methods are mathematically described below.

Least Squares

Let Y_1,..,Y_n be independent random variables with expected values μ_1,..,μ_n respectively and $B = [\beta_1, ..., \beta_m]^T$ be the parameter vector we want to estimate. Let us assume that expected values μ_i(i = 1...n) are functions of the parameter vector we want to estimate, then

$$E(Y_i) = \mu_i(\beta) \tag{2.1}$$

In the method of least squares, we want to find an estimator $\hat{\beta}$ that minimizes the sum of squares of the difference between Y_i and its expected value μ_i which is denoted as follows,

$$S = \sum [Y_i - \mu_i(\beta)]^2 \tag{2.2}$$

In case when random independent variables have unequal variances σ_i^2 we want to minimize the weighted sum of squared difference then equation 2.2 is changed as follows,

$$S = \sum w_i [Y_i - \mu_i(\beta)]^2 \tag{2.3}$$

where $w_i = (\sigma_i^2)^{-1}$. By using weights we ensure that observations with less reliability have less influence on the estimates.

Maximum likelihood

Let $Y = [y_1, ..., y_n]^T$ denote a random vector and let the joint probability density function be $f(Y;\Theta)$ which depends on the parameter vector $\Theta = [\theta_1, ..., \theta_m]^T$. Algebraically, the joint probability density function $f(Y;\Theta)$ is same as *likelihood function* $L(\Theta;Y)$. Let Ω denote the set of all possible values of the parameter vector Θ which is known as *parameter space*. The maximum likelihood estimator of θ is the value $\hat{\Theta}$ which maximizes the likelihood function which is,

$$L(\hat{\Theta}, Y) \geq L(\Theta, Y) \quad \forall \Theta \in \Omega \tag{2.4}$$

Often numerical methods and calculus are used to obtain the parameter estimates that maximize the likelihood or minimize the sum of squares.

2.1.4 Model Checking

Model checking is the process of evaluating the performance of a statistical model to ensure that it is a good fit for the data. This process typically involves examining the residuals of the model to see if they meet certain statistical criteria, such as being normally distributed and homoscedastic(constant variance) in nature. Let us consider Y_i set of response variables modelled as,

$$E(Y_i) = \mu_i; Y_i \sim N(\mu_i, \sigma^2)$$

Then the fitted values are the estimates $\hat{\mu_i}$ and residuals will be $y_i - \hat{\mu_i}$. In order to compare residuals of different observations we define standardized residuals as,

$$r_i = (y_i - \hat{\mu_i})/\hat{\sigma} \tag{2.5}$$

where $\hat{\sigma}$ is an estimate of the parameter σ.
The sum of squared residuals $\sum(y_i - \hat{\mu_i})^2$, which is a component that is optimized in parameter estimation is used to check the model adequacy.

2.1.5 Inference

Inference comes into play when we want to draw conclusions or make predictions about a process or population. Two main tools used in statistical inference are confidence intervals and hypothesis tests.
The next sections in this chapter introduce the data we used and the analysis we carried out to create a model. This step helps us to make informed decisions based on the data collected. In the below some concepts used during inference process are described.

2.2 Linear regression models

2.2.1 Description

In a simple linear regression model we want to model the relationship between two variable, for example, inflation and housing costs or literacy and income. Let us assume there is linear relationship between y and x. Then this relationship can be modelled as,

$$y = \beta_0 + \beta_1 x + \epsilon \tag{2.6}$$

where y is the dependent or response variable and x is the independent variable or predictor. Random variable ϵ is considered as the error term that is arising due to some external noise, β_0 is a constant and β_1 is a coefficient. Often times response variable y is influenced by more than one predictor. In this case linear relationship is defined as,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta n x_n + \epsilon \tag{2.7}$$

The parameters β_0,β_1,...,β_n are known as regression coefficients. ϵ brings in some randomness in y that is not explained the variables x. Here, relationship linearity is with parameters β and it is not necessarily linear with x variables. Now, let us consider a simple linear regression model for n observations represented as,

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i; i = 1, 2, ..., n \tag{2.8}$$

Here, we will assume that y_i and ϵ_i are random variables and x_i are known fixed values. We make the following assumptions for model in equation(refer the equation number),

1. $E(y_i) = \beta_0 + \beta_1 x_i$
2. var(y_i=σ^2)
3. cov(y_i,y_j)=0

First assumption implies that y_i depends only on x_i and all other variation is random. Second assumptions implies constant variance that is variance of y or ϵ does not depend on x_i values. Third assumption asserts that variables y are uncorrelated with each other.

2.2.2 Parameter estimation for linear model

As mentioned, we have known observations x_1,x_2,...,x_n and let us consider n random observations y_1,y_2,...,y_n then we can estimate parameters β_0 and β_1. Estimates $\hat{\beta}_0$, $\hat{\beta}_1$ are obtained using least squares method. In method of least squares we are interested to find $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squares of the deviations y_i - $\hat{y}_i$ of the n observed y_i's from their predicted values $\hat{y}_i$=$\hat{\beta}_0$+$\hat{\beta}_1 x_i$;

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{2.9}$$

To find $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the equation 2.9 we need to differentiate it with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and equating to zero leaves us with two equations,

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2.10}$$

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0 \tag{2.11}$$

Solution to equations 2.10 and 2.11 is given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{2.12}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.13}$$

2.2.3 Inference for linear model

Hypotheses about β_1 is important that hypotheses about β_0 since we are looking to find if some linear relationship exists between y and x. Here we will consider null hypothesis H_0: β=0 meaning there is no linear relationship between y and x in the model 2.8. In order to test for H_0: β=0 we assume that $y_i \sim N(\beta_0$+$\beta_1 x_i$,$\sigma^2)$. Here let us denote unbiased estimator of variance σ^2 i.e, s^2 as,

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} \tag{2.14}$$

Then $\hat{\beta}_1$ and s^2 have the properties,

1. $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2})$
2. $(n-2)s^2/\sigma^2 \sim \chi^2(n-2)$
3. $\hat{\beta}_1$ and s^2 are independent

From the above three properties we will have a score statistic t as,

$$t = \frac{\hat{\beta}_1}{s/\sqrt{\sum_i (x_i - \bar{x})^2}} \tag{2.15}$$

Using t we can determine significance and find out confidence intervals. All notations and expressions in section 2.2 are presented with reference to Rencher and Schaalje.

2.3 Effect size

Sometimes inferring results from a experiment especially in psychology should be done with measures of magnitude to understand the effect of the study because p-values in significance tests very much depend on sample sizes. For this we need effect size as a quantitative measure of study's effect. Larger effect size implies stronger effect and smaller effect size implies weaker effect.

2.3.1 Cohen's d

There are various measures of effect size but only Cohen's d measure is discussed as it relevant and useful in our work. Cohen's d is computed as,

$$d = (\mu_2 - \mu_1)/\sigma_g \tag{2.16}$$

Where μ_1, μ_2 are means of group 1 and group 2 respectively. σ_g is calculated as $\sqrt{\frac{(\sigma_1^2+\sigma_2^2)}{2}}$

- if d value is less than 0.2 then we say effect is small.
- if d value is around 0.5 then we say effect is medium.
- if d value is around 0.8 then we say effect is large.

3. Experiment

We designed an experiment in which we wanted to understand if the tracking accuracy of the participants is improved when we modify low-scoring trajectories obtained from our analysis according to the two metrics we selected. In the following design section modification made to the original trajectories is explained in detail. In later sections of this chapter, we describe the apparatus used, the stimuli that are presented, and the experimental procedure carried out.

3.1 Data

As mentioned in the introduction we used data from the work of Dechterenko (2017) in which they investigated if flipping objects during the trial both horizontally and vertically can be used as a masking technique in long and repetitive tasks so that participants would not notice repeated trials. The data from this experiment consisted of trajectory data that were present during trials and response data collected. Based on the column names shown in table 2.1 trajectory data

TrackId	X1	Y1	X2	Y2	X3	Y3	X4	Y4	X5	Y5	X6	Y6	X7	Y7	X8	Y8	Time
1	10.51	1.09	-1.09	-5.21	-10.04	-6.80	-6.18	3.10	-10.36	3.64	-9.23	-3.98	-7.14	7.27	-11.93	3.68	0.01
1	10.52	1.15	-1.15	-5.22	-9.98	-6.80	-6.20	3.04	-10.31	3.61	-9.22	-4.04	-7.19	7.24	-11.89	3.65	0.02
1	10.53	1.20	-1.21	-5.22	-9.92	-6.81	-6.21	2.99	-10.25	3.59	-9.20	-4.09	-7.24	7.22	-11.83	3.63	0.04
1	10.53	1.26	-1.27	-5.21	-9.86	-6.82	-6.24	2.93	-10.20	3.57	-9.17	-4.14	-7.30	7.20	-11.78	3.60	0.05
1	10.53	1.32	-1.33	-5.21	-9.81	-6.81	-6.25	2.88	-10.14	3.56	-9.12	-4.18	-7.35	7.19	-11.73	3.58	0.06
1	10.54	1.38	-1.39	-5.21	-9.75	-6.81	-6.25	2.82	-10.08	3.55	-9.08	-4.22	-7.41	7.17	-11.67	3.55	0.07

Table 3.1: Sample of trajectory data

contains a column called *TrackId* which is an identity given to each trajectory that is generated. We need to have a trajectory ID because during the analysis of results, we can identify on which trajectories participants did not perform well and study characteristics of those low-scoring trajectories. Columns X_i,Y_i(where $i = 1, 2, 3, 4$) contain coordinates of 4 target objects and columns X_j, Y_j(where $j = 5, 6, 7, 8$) contain coordinates of 4 distractor objects. Values in column *Time* represent each timestep of the trial, In the experiment from which this data is obtained each trial lasted for 6 seconds with each timestep being 85th of a second. MOT being a dynamic task it seems obvious to have a time column but apart from that sometimes choosing timesteps could affect the experience of the participants. In this data each datapoint is a frame with eight objects shown at that particular timestep, each trajectory in this data contains 510(6*85) datapoints and a total of 500 trajectories. Please note that due to space constraint, floating values in all tables are rounded to 2 decimals.

Table 2.2 shows a sample of response data after filtering columns that are not useful in the analysis. Response data also contains column *TrackId* and its description is the same as mentioned above. The second column *nCorrect* gives how many target objects the participant tracked correctly and the values in the third column *accuracy* reflect the number of target objects tracked in *nCorrect*. Each data point in response data gives performance on a particular trajectory in a specific trial. The number of data points in the response data depends on the number of trials that are presented in the experiment. Out of 500 trajectories

TrackId	nCorrect	accuracy
173	3	0.75
38	1	0.25
167	4	1.00
126	4	1.00
12	3	0.75
14	3	0.75

Table 3.2: Sample of response data after filtering unwanted columns

in trajectory data 450 trajectories were used in the experiment so response data contains 450 unique *TrackId's* and the remaining 21 trajectories are excluded for further analysis.

3.2 Analysis

Initially based on the accuracy values in the response data we identified some trajectories where participants could not track all target objects when these trajectories were presented to them. After identifying trajectories we came up with a couple of descriptors that would quantify the difficulty of the trajectory like crowding, some centroid measures, mean area of the convex hull formed by the target objects. Apart from the mentioned, we tried other metrics like the number of targets in each hemifield and the intersection area when convex hulls formed by targets and distractors overlap but in order to keep the model simple and avoid collinearity we didn't consider some metrics for the model. These measures we tried are basically related to previous works mentioned in chapter 1. After summarizing data in table 2.2 by calculating the mean accuracy of each trajectory the data looks in table 2.3 and here we removed *nCorrect* as it no longer needed because we can work just with *accuracy*.

TrackId	accuracy
442	0.95
486	0.92
244	0.75
79	0.88
470	0.25
99	0.88

Table 3.3: Summarized response data

At this point, we excluded all trajectories with mean accuracy 1 because participants could track all the targets on those trajectories correctly and after exclusion we had 199 trajectories in total. These 199 trajectories are the ones where one or more participants could not track all the targets successfully.
When computing the metrics we would like to determine some statistics particular to target objects or distractor objects for this reason it is hard to work with data in table 3.1 as it is so we transform the trajectory data to long format which looks like,

TrackId	Time	object	X	Y	target
1	0.01	1	10.51	1.09	TRUE
1	0.01	2	-1.09	-5.21	TRUE
1	0.01	3	-10.04	-6.80	TRUE
1	0.01	4	-6.18	3.10	TRUE
1	0.01	5	-10.36	3.64	FALSE
1	0.01	6	-9.23	-3.98	FALSE
1	0.01	7	-7.14	7.27	FALSE
1	0.01	8	-11.93	3.68	FALSE
1	0.02	1	10.52	1.15	TRUE
1	0.02	2	-1.15	-5.22	TRUE
1	0.02	3	-9.98	-6.80	TRUE
1	0.02	4	-6.20	3.04	TRUE
1	0.02	5	-10.31	3.61	FALSE
1	0.02	6	-9.22	-4.04	FALSE
1	0.02	7	-7.19	7.24	FALSE
1	0.02	8	-11.89	3.65	FALSE

Table 3.4: Sample of trajectory data in long format

After summarizing accuracy of each trajectory we calculated 6 metrics and summarized them for each trajectory. Summarized data for each trajectory looks as in table 2.4 and actual data in table 2.4 consists of summary results for 93 trajectories.

TrackId	accuracy	Metric1	Metric2	Metric3	Metric4	Metric5	Metric6
3	0.67	8.66	8.50	1.28	13.77	66.06	5.31
18	0.96	7.09	6.58	2.20	11.36	67.32	5.36
19	0.50	9.19	7.94	1.96	16.13	88.98	6.75
23	0.96	8.48	8.01	2.12	14.76	107.37	8.61
27	0.75	9.56	9.40	4.23	13.95	150.02	8.66
28	0.97	8.69	6.98	0.27	12.76	50.83	9.98

Table 3.5: Merged data showing a summary of accuracy and metrics for each trajectory

Next, we performed bivariate analysis to see how accuracy is influenced by each metric and the results of this analysis can be seen in figure 2.1. Looking at the regression line in each plot in figure 2.1 we can find out some impact of metric 5 and metric 6 on accuracy but others metrics seems to explain nothing from the plots.

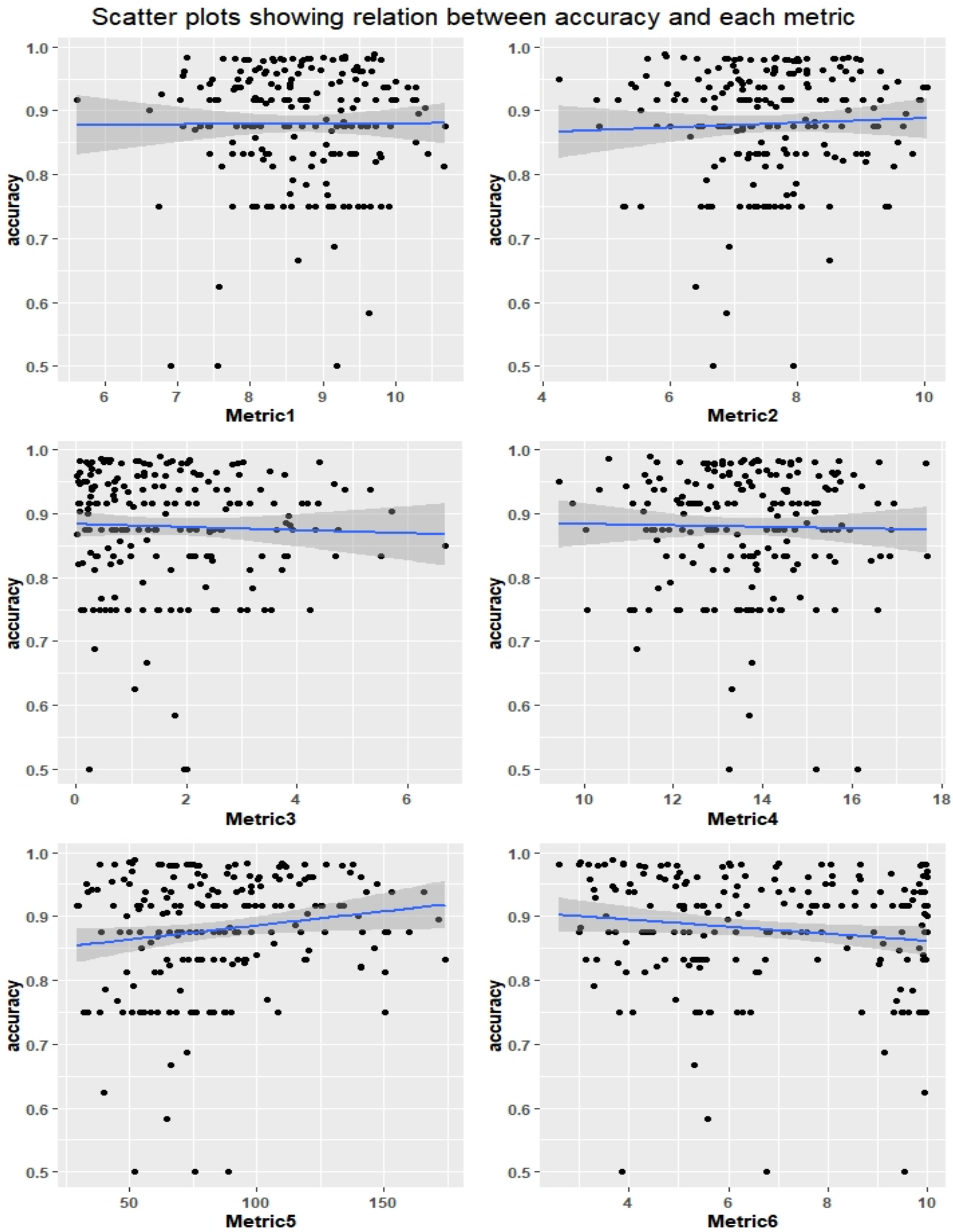

Figure 3.1: Scatterplots generated as part of bivariate analysis

The description of each metric is as follows,

- Metric 1 - Mean distance of target objects to the centroid formed by all objects i.e, targets and distractors.
- Metric 2 - Mean distance of target objects to the centroid formed by only targets
- Metric 3 - Minimum distance of a target from the centroid formed by only targets
- Metric 4 - Maximum distance of a target from the centroid formed by only targets
- Metric 5 - Mean area of convex hull formed by targets
- Metric 6 - Maximum crowding(assuming gaze point at objects centroid)

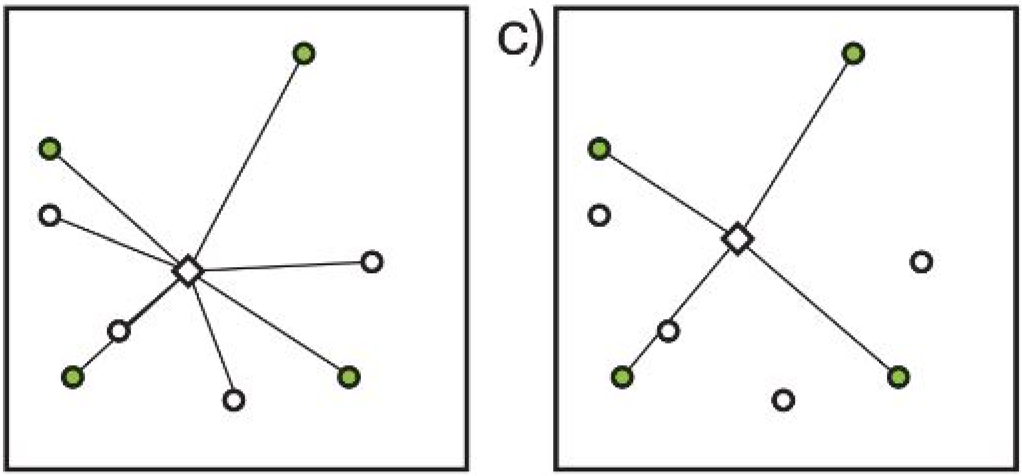

Figure 3.2: Instance showing object positions in MOT[Source:Lukavskỳ]

Metric 1 can be understood from figure 3.2(left) the diamond symbol is centroid formed by all the objects and lines from each target(green dots) indicates to centroid can be considered as distance to it and we calculate mean of four distances. Metrics 2,3 and 4 can be understood from figure 3.2(right) where diamond symbol represents centroid formed by targets and lines indicate distance of each target to the centroid. We calculate mean,minimum and maximum of these distances. Metric 6 is computed with equation 1.2 with gaze at all object centroid and we will take maximum of the computed result.

After bivariate analysis, we created a linear models with accuracy as the response variable and each metric as explanatory variable this is discussed in the following section.

3.3 Models

In this section we present the summary of 6 linear models we tested based the six metrics that are mentioned above as part of bivariate analysis. Then based on the model summary results we choose to pick two models that will be tested experimentally on new set of participants to see how their tracking performance varied based on the predictor. Please note that all models are fitted on trajectory descriptors or metrics obtained from all 199 trajectories.

Table 3.6: Summary of model 1 with metric 1 as predictor

	Dependent variable:
	accuracy
Metric1	0.001 (0.007)
	t = 0.072
	p = 0.943
Constant	0.875 (0.064)
	t = 13.604
	p = 0.000***
Observations	199
R^2	0.00003
Adjusted R^2	−0.005
Residual Std. Error	0.093 (df = 197)
F Statistic	0.005 (df = 1; 197)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3.7: Summary of model 2 with metric 2 as predictor

	Dependent variable:
	accuracy
Metric2	0.004 (0.006)
	t = 0.620
	p = 0.536
Constant	0.852 (0.045)
	t = 18.854
	p = 0.000***
Observations	199
R^2	0.002
Adjusted R^2	−0.003
Residual Std. Error	0.093 (df = 197)
F Statistic	0.384 (df = 1; 197)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3.8: Summary of model 3 with metric 3 as predictor

	Dependent variable:
	accuracy
Metric3	−0.002 (0.005)
	t = −0.504
	p = 0.616
Constant	0.883 (0.010)
	t = 84.902
	p = 0.000***
Observations	199
R^2	0.001
Adjusted R^2	−0.004
Residual Std. Error	0.093 (df = 197)
F Statistic	0.254 (df = 1; 197)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3.9: Summary of model 4 with metric 4 as predictor

	Dependent variable:
	accuracy
Metric4	−0.001 (0.004)
	t = −0.275
	p = 0.784
Constant	0.895 (0.058)
	t = 15.517
	p = 0.000***
Observations	199
R^2	0.0004
Adjusted R^2	−0.005
Residual Std. Error	0.093 (df = 197)
F Statistic	0.076 (df = 1; 197)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3.10: Summary of model 5 with metric 5 as predictor

	Dependent variable:
	accuracy
Metric5	0.0004 (0.0002)
	t = 2.163
	p = 0.032**
Constant	0.841 (0.019)
	t = 45.188
	p = 0.000***
Observations	199
R^2	0.023
Adjusted R^2	0.018
Residual Std. Error	0.092 (df = 197)
F Statistic	4.677** (df = 1; 197)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3.11: Summary of model 6 with metric 6 as predictor

	Dependent variable:
	accuracy
Metric6	−0.006 (0.003)
	t = −1.944
	p = 0.054*
Constant	0.917 (0.020)
	t = 44.834
	p = 0.000***
Observations	199
R^2	0.019
Adjusted R^2	0.014
Residual Std. Error	0.092 (df = 197)
F Statistic	3.778* (df = 1; 197)
Note:	*p<0.1; **p<0.05; ***p<0.01

There are different approaches in selecting predictors for a regression model that captures the relationship with dependent variable in the best possible way. Usually, predictors with highest adjusted R^2 are selected and predictors with p-values greater than 0.05 are disregarded. For our work we considered both Adjusted R^2 and p-values to select two models. Looking at the summary of each model, model 5 with mean area of convex hull as predictor(metric 5) should be one of the model to be selected due to its highest adjusted R^2 and least p-value and model 6 with maximum crowding as predictor should be the other model selected as per its adjusted R^2 and p-values. However, we did not use model 5 due to its

complexity and limited property that is the chances of convex hulls not formed or do not remain intact is high. Modifying trajectories(details described in section 3.4) based on convex hull is slightly harder compared modifying trajectories just based on distance measures. For our experiment, we selected model 6 because crowding is a general phenomenon it is observed in many cases and we selected model 4 whose predictor is maximum distance between a target to targets centroid to see if by minimizing the distance to target centroid would improve tracking accuracy(minimizing the maximum distance).
Originally, our idea is to select two sets of trajectories from 199 trajectories again based on accuracy. For this purpose, we selected one set of trajectories whose accuracy values are under the first quartile(Q_1) and another set of trajectories whose accuracy values are above third quartile(Q_3). Filtering 199 trajectories by

```
responses = read.csv("responses.txt",head=TRUE,sep=",")

summary(responses$accuracy)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.5000  0.8333  0.9062  0.8792  0.9482  0.9886
```

Figure 3.3: summary of accuracy showing quartiles,central measure and min-max

accuracy ≥ 0.9482 and accuracy ≤ 0.8333 we got 93 trajectories in total. From 93 trajectories we selected 16 trajectories whose accuracy is under Q_1 and other 16 trajectories whose accuracy is above Q_3. We modified 16 trajectories according to metric 6(crowding) and 16 trajectories according to metric 4(maximum distance of target to targets-centroid). This modification procedure is explained in the following design section.

3.4 Design

Our main goal while designing the experiment is to come up with a procedure that would detect the peak and suppress the peak. Here peak can be interpreted as the most difficult part of tracking a trajectory. After identifying the low-scoring trajectories, we modified trajectories based on two metrics which are crowding(Metric A) and the maximum distance of target objects to the targets-centroid(Metric B) that we selected to quantify the difficulty of the trajectory.
we tried to modify the original trajectory as follows,

Step 1: For each trajectory, the metric quantifying difficulty is calculated
Step 2: Identify a time interval that contains the maximum value of the metric
Step 3: Perform linear interpolation on the original trajectory data within the time interval
Step 4: Calculate midpoints between the interpolated part of the trajectory and its original points
Step 5: Replace original trajectory points in the time interval at which peak values are detected with the midpoints calculated so that we will have a modified trajectory with a smaller peak value.

In the procedure described above, we used interpolation technique to suppress the peak but we cannot use interpolated points for the stimuli because it makes the tracking task quite easy and participants can predict the trajectory's movement during the trial. In order to overcome this we calculated the midpoints between interpolated points and original trajectory points in the peak interval. In MOT tasks the object directions are generated from a random distribution like von Mises distribution(Chapter 45 in Forbes et al.), by calculating the midpoints we could maintain the movement of the objects in the modified trajectory still in a random way so that participants would not be able to predict future locations of the objects.
The design procedure is explained with the help of two trajectories with IDs 90, 409 that we used in our work. Metric 1 manipulations are explained using trajectory-90 and metric 2 manipulations are explained using trajectory-409 in the following subsections.

3.4.1 Metric A - Crowding

In regards to metric A, we tried to manipulate the inter-object spacing of all 4 targets based on the crowding values of the original trajectory. Figure 3.3 shows the position of objects at the peak crowding for the original trajectory(on the left) and modified trajectory(on right). Inter-object spacing manipulation done using the modification procedure explained is reflected within squared and circled regions in figure3.3

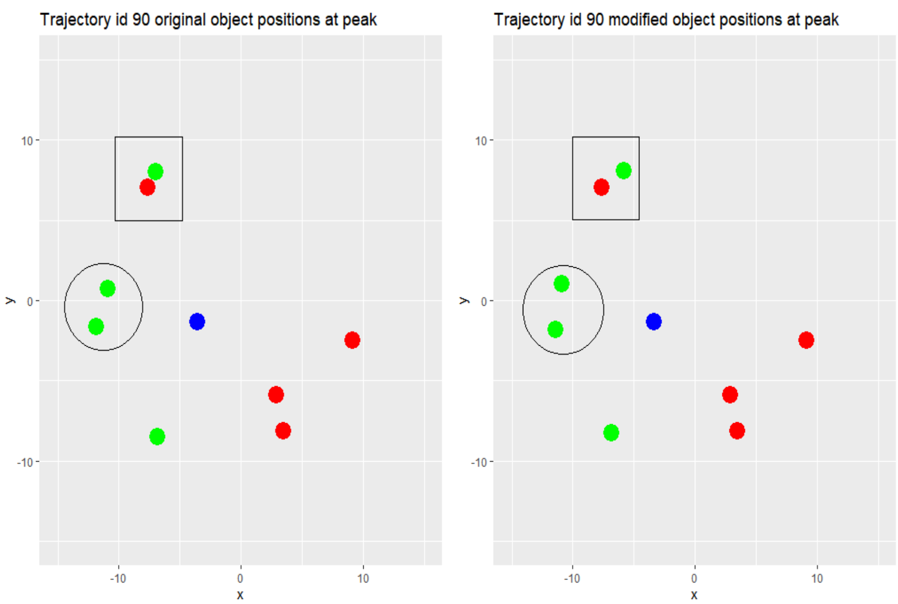

Figure 3.4: Snapshot showing positions of the objects for original trajectory and its modified version at the peak value of metric A. In the figure green, red, and blue dots represent targets, distractors and all objects centroid respectively

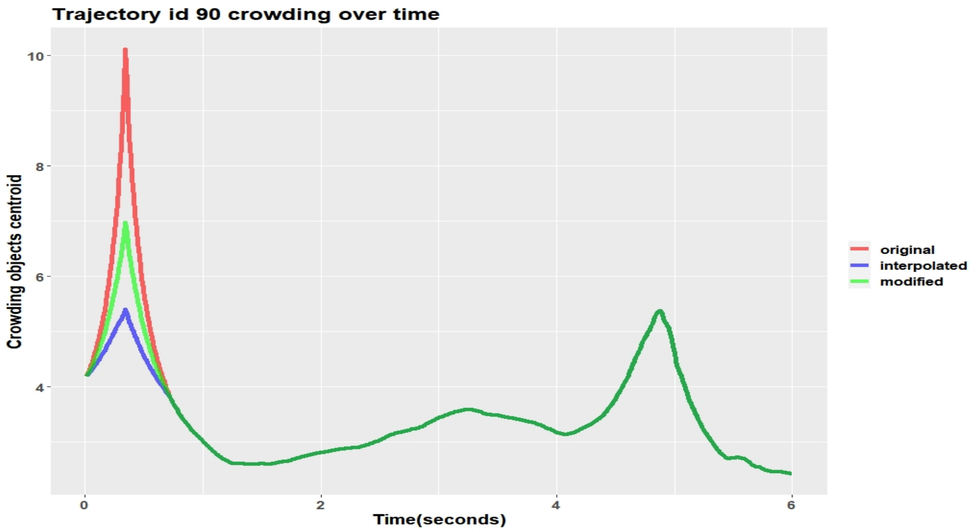

Figure 3.5: Plot showing how crowding varies over time for the original trajectory and its modified version

Figure 3.4 shows an output of the modification process with Metric 1. Crowding values were calculated for both the original(red) and modified (green)trajectory. In this figure, we can notice that the peak crowding value of the modified trajectory is smaller than the crowding value of its original trajectory.

3.4.2 Metric B - Maximum distance of the target from targets centroid

For metric B we modified only the position of the target object that is at a maximum distance from the centroid formed by all target objects. Figure 3.5, shows the position of the objects at timestep for a trajectory with ID 409 where the distance of the target object to its centroid is maximum. In figure 3.5 target object in green marked inside a square is at the maximum distance from the centroid of the targets (blue point marked inside a triangle) and one can notice the reduced distance in the right side plot of figure3.5

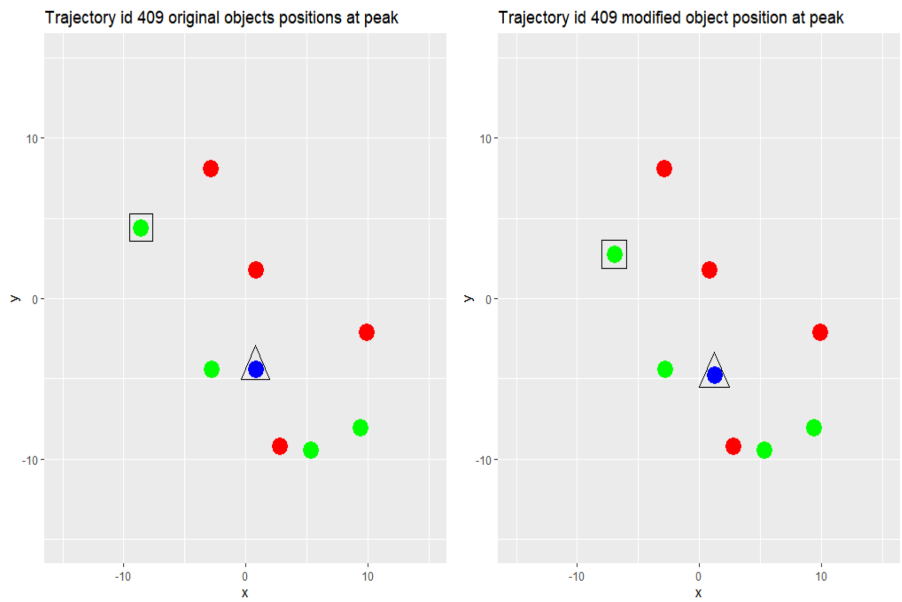

Figure 3.6: Snapshot showing positions of the objects for original trajectory and its modified version at the peak value of metric B. In the figure green, red, and blue dots represent targets, distractors, and targets centroid respectively

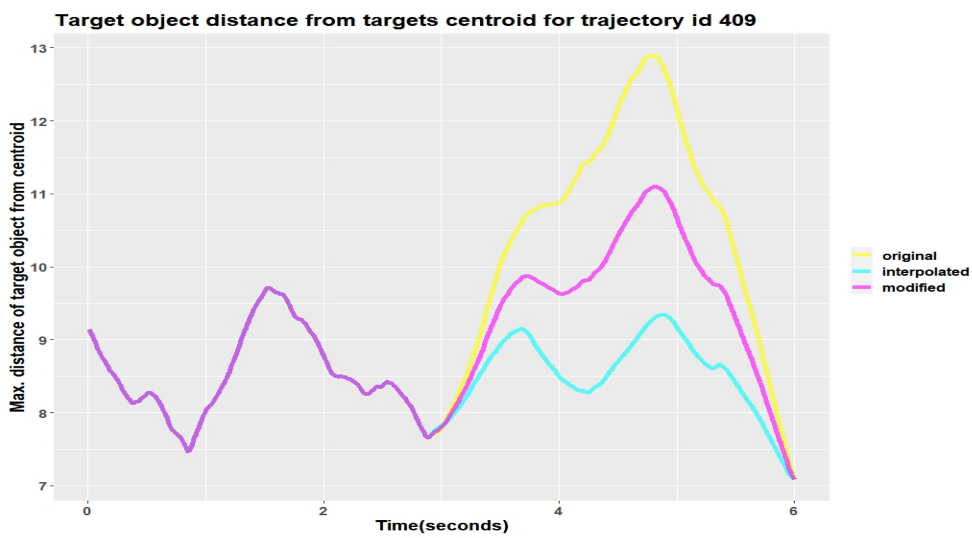

Figure 3.7: Plot showing how the distance of a target to targets centroid over time for the original and its modified trajectory

Figure 3.6 shows the output of metric 2 values calculated for both the original(yellow) and modified (magenta)trajectory. In this figure, we can notice that the distance between the target object and the centroid of the target is decreased for the modified trajectory compared to the original.

3.5 Method

3.5.1 Participants

A total of 75 students consisting of 48 females and 27 males participated in the experiment, mean age of the participants was 21.44 years with ages ranging from 19 to 28 years. All participants had normal or corrected vision and voluntarily took part in the experiment.

3.5.2 Apparatus and stimuli

This experiment was programmed in Python using PsychoPy software Peirce et al. (2022) and presented on a 22” LCD screen (1920 x 1080 px). The participant sat approximately 50 cm from the screen, and the head position was not controlled. Stimuli used in this experiment include eight white objects 1° against grey background. Each of them moved with a constant speed of 5°. All objects moved in 30° diameter inside circular arena. Objects bounced off the circular border back to screen center with a random change to avoid prediction of motion paths by the participant. Objects also bounced off each other allowing atleast 0.1° space between them. The direction of the objects of original trajectories(not modified) was sampled from a von Mises distribution with a concentration parameter of k=40 for each frame(Děchtěrenko et al. (2017)).

3.5.3 Procedure

The experiment consisted of 64 trials and an extra 6 trials are presented for practice before the experiment. In each trial, eight objects are presented at random positions on the screen and target objects are cued by flashing. After targets are cued all 8 objects move for 6 s and stop. Once objects stop moving participants are asked to click on the objects they tracked. Out of 64 trials, 32 trials are modified and the remaining 32 trials are the original version of the modified trials. In this experiment, all participants are presented with the same trajectories but in a different order based on the protocol file selected. In presenting the trials we made sure that trials consisting of modified trajectory and it’s original version are at least 16 trials apart so that participants do not notice the trajectories. A sample of how a protocol file created for experimental trials is shown in table 3.12

prot.id	trial.id	trajectory.id	manipulated	metric	performance
22	1	409	yes	metric.B	Q1
22	2	495	yes	metric.B	Q1
22	3	173	yes	metric.B	Q1
22	4	276	yes	metric.B	Q1
22	5	283	yes	metric.B	Q3
22	6	200	yes	metric.B	Q3
22	7	102	yes	metric.B	Q3
22	8	447	yes	metric.B	Q3
22	9	450	yes	metric.B	Q3
22	10	398	yes	metric.B	Q3

Table 3.12: A sample of protocol file

In every protocol file three main columns that it should contain are a column with protocol ID to identify each protocol created, a column with trial ID given to each trial and third column containing stimulus ID given to each stimulus. In the table columns 'prot.id', 'trial.id' and 'trajectory.id' contains protocol ID, trial ID and stimulus ID respectively. The remaining columns in the protocol file can be set as per the experiment requirements as in if the experimenter wants to control some aspect of the stimulus those parameters can be listed in the file. In our experiment we used information in column 'manipulated' to present either the original trajectory or the modified trajectory.

4. Results

In this chapter we present results obtained from data collected in our experiment. Firstly, to have an idea of how collected data looks like check table 4.1

prot_id	trial_id	trajectory_id	participant_id	modified	metric	performance	mouse.clicked_name	nCorrect
15	45	382	35	no	metric.A	Q3	['o1_copy_3', 'o1_copy_0', 'o1_copy_1', 'o1_copy_2']	4.00
2	61	118	22	yes	metric.A	Q3	['o1_copy_4', 'o1_copy_3', 'o1_copy_0', 'o1_copy_2']	3.00
1	56	126	41	yes	metric.B	Q3	['o1_copy_1', 'o1_copy_2', 'o1_copy_4', 'o1_copy_7']	2.00
19	63	346	19	yes	metric.B	Q1	['o1_copy_2', 'o1_copy_1', 'o1_copy_0', 'o1_copy_3']	4.00
15	20	102	35	no	metric.B	Q3	['o1_copy_1', 'o1_copy_0', 'o1_copy_2', 'o1_copy_3']	4.00
7	34	118	7	no	metric.A	Q3	['o1_copy_2', 'o1_copy_0', 'o1_copy_3', 'o1_copy_1']	4.00
4	25	382	4	no	metric.A	Q3	['o1_copy_1', 'o1_copy_2', 'o1_copy_3', 'o1_copy_0']	4.00
12	3	283	12	yes	metric.B	Q3	['o1_copy_2', 'o1_copy_0', 'o1_copy_3', 'o1_copy_1']	4.00
14	53	415	54	yes	metric.A	Q1	['o1_copy_1', 'o1_copy_0', 'o1_copy_2', 'o1_copy_3']	4.00
2	7	496	22	yes	metric.B	Q1	['o1_copy_3', 'o1_copy_2', 'o1_copy_1', 'o1_copy_0']	4.00

Table 4.1: Sample of processed data collected from experiment

Most of the columns by what they mean in table 4.1 are explained previously except for one column "mouse.clicked_name" and what this column captures a list of four objects participant selected by mouse clicks after end of a trial. The set of objects ['o1_copy_0', 'o1_copy_1', 'o1_copy_2', 'o1_copy_3'] corresponds to target objects and ['o1_copy_4', 'o1_copy_5', 'o1_copy_6', 'o1_copy_7'] corresponds to distractors. Data in this column will be transformed as "nCorrect" to get on number of correctly tracked objects and once we have this column "mouse.clicked_name" can be ignored for further analysis.

In the plots above, each point represents the mean tracking accuracy of each participant on both original trajectories and modified trajectories for the two metrics. Green points show the mean accuracy of participants on original trajectories, blue points show the mean accuracy on their modified ones, and black points represent the mean of all the participants. The pair of points connected by the lines correspond to each participant. Looking at the plots we can observe that the mean accuracy of all participants is slightly higher on the modified trajectories for both metrics especially in case of Metric.A . However, we need to test if this difference is significant in the case of both metrics. Other observations that we can make from looking at the connecting lines is that there are multiple cases where some participants did well on modified trajectories compared to their original ones and there are also cases where some participants performed well on the original trajectories and did not do well on their modified ones. Please note that in figure 4.1 there is high overlapping of the points as many values are very close. So, in order to have a better view of the plot the point positions have been jittered horizontally due to this one can notice some mean accuracy values are above 1.0 but in actual data, all values are equal to 1.0 or less than 1.0
For further analysis, for each participant we calculated the difference in accuracy between the modified trajectory and the original trajectory and then we computed the average of this difference for each metric A and metric B separately and results shown in figure 4.2.

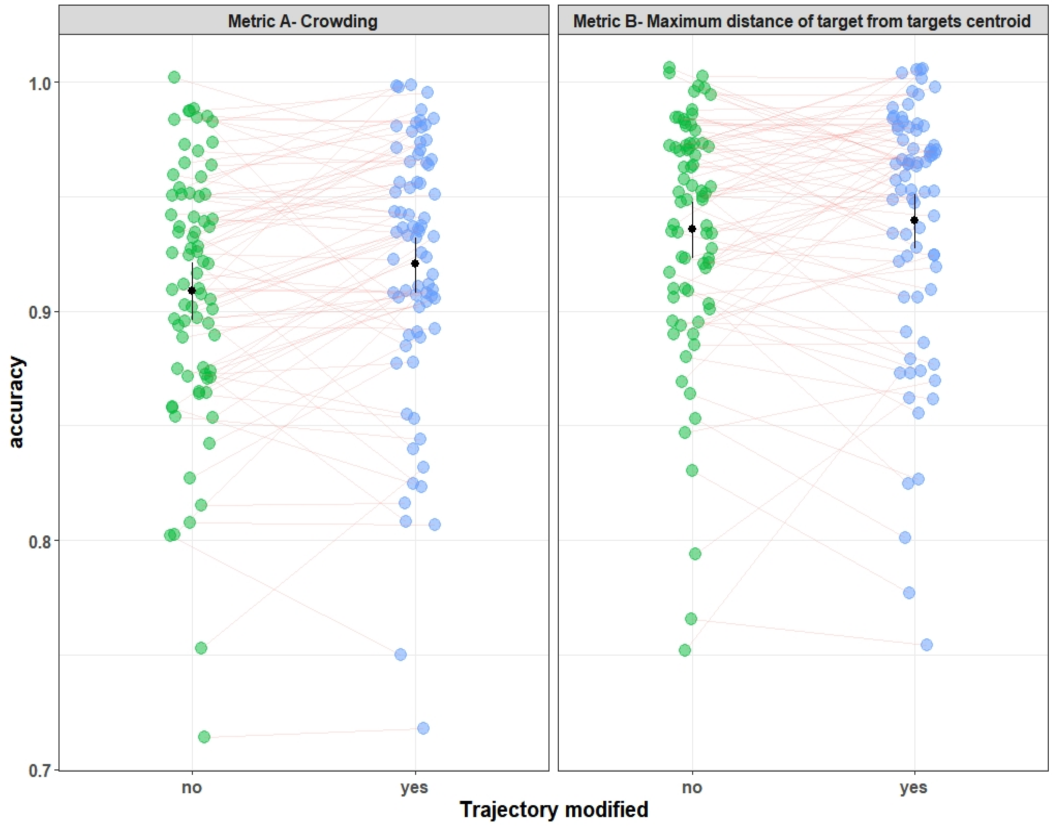

Figure 4.1: Plot showing mean accuracy of each participant on both original and modified trajectories of each metric

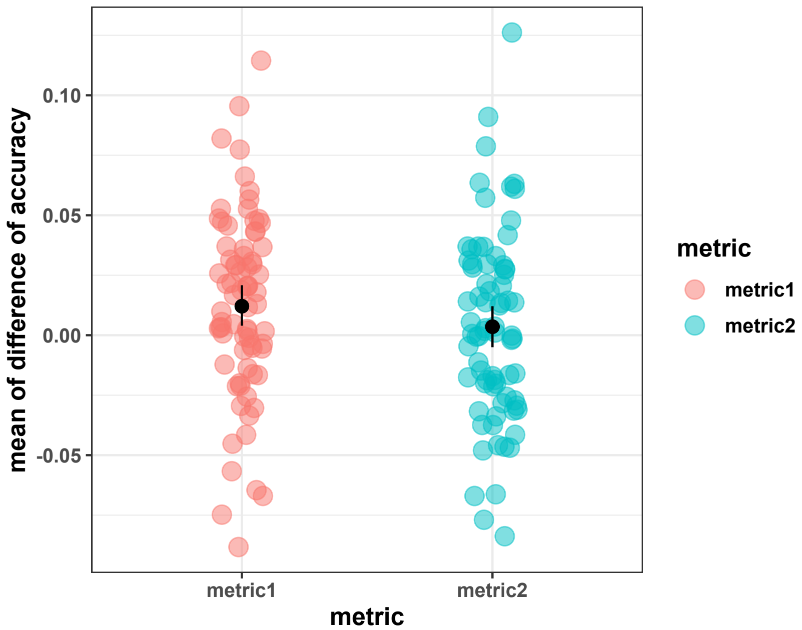

Figure 4.2: plot showing difference of accuracy between original and modified trajectories averaged for each participant. Here metric 1 and metric 2 in figure means metric.A and metric.B respectively

In order to understand how significant the relationship is between dependent and explanatory variables we did a T-test. From T-test results are t=1.3753, p<.05 and to say t is relatively large but it is not that large to ignore some relationship between tested metrics and accuracy.

	Cohens_d	CI	CI_low	CI_high
Metric.A	0.32	0.95	0.08	0.55
Metric.B	0.09	0.95	-0.14	0.32

Table 4.2: Effect sizes of both metrics

In order to understand the effect of each metric.A, metric.B on participants tracking accuracy we computed cohen's d individually and those results are mentioned in table 4.2. For metric.A cohen's d is 0.32 that means there is some moderate effect on tracking accuracy whereas, for metric.B cohen's d is 0.09 meaning there is effect of metric on the tracking accuracy.

5. Discussion

In this study we came up with two predictive models with an expectation that these model could explain how the tracking performance varies in an MOT experiment. For validating our models we used both original trajectories at hand and modified trajectories this approach can be consider as analogy to experimental group and control group during clinical trials of testing some medicine. In experimental group the subjects receive some treatment and researcher is interested to find out the effect whereas in control group subjects receive no treatment and their effect is already seen. In the same, we modified trajectories to see the effect on tracking performance with an assumption that it would improve and original trajectories are also presented for comparison.

5.1 Model 1

In model 1 the predictor variable we used to assess tracking performance is maximum crowding. We considered this because crowding is a well studied phenomenon and previous research has shown that this effect would deplete tracking performance in MOT tasks. Based on our results it is understand that minimizing crowding of the original trajectories had a moderate effect on improving the tracking accuracy.

5.2 Model 2

In model 2 the predictor variable we used to explain tracking accuracy is a distance measure calculated as distance between one of the target objects and centroid formed by the targets whose distance is the farthest. The reason for considering this measure to quantify difficulty of trajectories is again from previous studies which used eye tracking device to capture eye movements during MOT tasks have identified that people use some of target-centroid looking strategy and imagine polygon with targets as vertices. So when a target is far from its centroid we assumed that there is high chance of losing that target during tracking due to some limitations in peripheral vision. We test this assumption by reducing the target object distance to its center and see if the tracking performance improved. However, there is no supporting evidence found and it effect is negligible.

Conclusion

In this thesis, we have used linear regression models to predict tracking accuracy in Multiple object tracking experiment. Firstly, we obtained some data collected from a previous MOT experiment and based on results of that experiment we have identified some trajectories where some participants could not track all the target correctly. After we have identified those trajectories we computed some trajectory descriptors that would quantify the tracking difficulty. Then we modified these trajectories with low tracking accuracy to make them more easier to track. We have tested two models and in one model where dificulty quantified by crowding explains improved tracking performance to a little extent and the other model where difficulty is quantified by distance measure related to only targets seems to have no impact in improving tracking accuracy. Our interpretation from results of both these models is that first model definitely creates a moderate effect because inter-object spacing of all objects both targets and distractors are considered and modified. Whereas, in model 2 only object spacing of one target was modified with respect to a centroid this could be the reason for negligible effect or may the object that was modified is not the toughest to track among the four objects. Definitely, in future work it is worthy to try more descriptors involving convex hulls and distance measures calculated by giving more preference to distractor objects. Practically MOT can be applied in evaluating visual attention of a sport person playing some team sport or assessing candidates who work in air traffic control so by doing these studies in understanding and predicting tracking accuracy there could be some benefit.

Bibliography

G. A. Alvarez and P. Cavanagh. Independent resources for attentional tracking in the left and right visual hemifields. *Psychological science*, 16(8):637–643, 2005.

G. A. Alvarez and S. L. Franconeri. How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of vision*, 7(13): 14–14, 2007.

E. Awh and H. Pashler. Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2):834, 2000.

H. Bouma. Interaction effects in parafoveal letter recognition. *Nature*, 226(5241): 177–178, 1970.

P. Cavanagh and G. A. Alvarez. Tracking multiple targets with multifocal attention. *Trends in cognitive sciences*, 9(7):349–354, 2005.

N. Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(1):87–114, 2001.

J. C. Culham, P. Cavanagh, and N. G. Kanwisher. Attention response functions: characterizing brain areas using fmri activation during parametric variations of attentional load. *Neuron*, 32(4):737–745, 2001.

F. Dechterenko and J. Lukavsky. Predicting eye movements in multiple object tracking using neural networks. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, pages 271–274, 2016.

F. Děchtěrenko, J. Lukavskỳ, and K. Holmqvist. Flipping the stimulus: Effects on scanpath coherence? *Behavior Research Methods*, 49(1):382–393, 2017.

A. J. Dobson and A. G. Barnett. *An introduction to generalized linear models*. Chapman and Hall/CRC, 2018.

H. M. Fehd and A. E. Seiffert. Eye movements during multiple object tracking: Where do participants look? *Cognition*, 108(1):201–209, 2008.

H. M. Fehd and A. E. Seiffert. Looking at the center of the targets helps multiple object tracking. *Journal of vision*, 10(4):19–19, 2010.

C. Forbes, M. Evans, N. Hastings, and B. Peacock. *Statistical distributions*, volume 4. Wiley New York, 2011.

S. Franconeri, S. Jonathan, and J. Scimeca. Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological science*, 21(7):920–925, 2010.

S. L. Franconeri, J. Y. Lin, J. T. Enns, Z. W. Pylyshyn, and B. Fisher. Evidence against a speed limit in multiple-object tracking. *Psychonomic bulletin & review*, 15(4):802–808, 2008.

J. E. Hoffman and B. Nelson. Spatial selectivity in visual search. *Perception & Psychophysics*, 30(3):283–290, 1981.

J. Intriligator and P. Cavanagh. The spatial resolution of visual attention. *Cognitive psychology*, 43(3):171–216, 2001.

J. Jovicich, R. J. Peters, C. Koch, J. Braun, L. Chang, and T. Ernst. Brain areas specific for attentional load in a motion-tracking task. *Journal of cognitive neuroscience*, 13(8):1048–1058, 2001.

D. M. Levi. Crowding—an essential bottleneck for object recognition: A mini-review. *Vision research*, 48(5):635–654, 2008.

J. Lukavskỳ. Eye movements in repeated multiple object tracking. *Journal of Vision*, 13(7):9–9, 2013.

H. S. Meyerhoff, F. Papenmeier, and M. Huff. Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, 79(5):1255–1274, 2017.

G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

M. M. Müller, P. Malinowski, T. Gruber, and S. Hillyard. Sustained division of the attentional spotlight. *Nature*, 424(6946):309–312, 2003.

L. Oksama and J. Hyönä. Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? an individual difference approach. *Visual cognition*, 11(5):631–671, 2004.

J. Peirce, R. Hirst, and M. MacAskill. Building experiments in psychopy (2nd editio), 2022.

Z. Pylyshyn. The role of location indexes in spatial perception: A sketch of the finst spatial-index model. *Cognition*, 32(1):65–97, 1989.

Z. W. Pylyshyn. Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1-2):127–158, 2001.

Z. W. Pylyshyn and R. W. Storm. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, 3(3):179–197, 1988.

A. C. Rencher and G. B. Schaalje. *Linear models in statistics*. John Wiley & Sons, 2008.

J. M. Scimeca and S. L. Franconeri. Selecting and tracking multiple objects. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(2):109–118, 2015.

J. T. Serences and S. Yantis. Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. *Cerebral cortex*, 17(2):284–293, 2007.

W. M. Shim, G. A. Alvarez, T. Vickery, and Y. Jiang. The number of attentional foci and their precision are dissociated in the posterior parietal cortex. *Cerebral cortex*, 20(6):1341–1349, 2010.

Y. Xu and M. M. Chun. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440(7080):91–95, 2006.

S. Yantis. Multielement visual tracking: Attention and perceptual organization. *Cognitive psychology*, 24(3):295–340, 1992.

G. J. Zelinsky and M. B. Neider. An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, 16(5):553–566, 2008.

List of Figures

List of Tables

List of Abbreviations

MOT - Multiple object tracking
fMRI - Functional magnetic resonance imaging
EEG - Electroencephalogram
LGN - Lateral geniculate nucleus
CNN - Convolutional neural network
EDA - Exploratory data analysis

A. Attachments

A.1 First Attachment

We programmed most of the project in R with the help Rstudio and for the experiment we used PsychoPy based on python programming language which has Graphical interface to set experiment parameters as well as coder view to program. The thesis attachment contains code, data, images, videos and experiment files.

A.1.1 Code

- install_load_packages.R - installs and loads all libraries used in the project
- load_preprocess_data.R - for loading and preprocessing trajectory and response data
- functions.R - contains all the functions implemented for the project running the script will load all functions to R environment in Rstudio
- bivariate_analysis.R - generates some trajectory statistics, prints scatter plots and creates linear models
- modify_metricA_tracks.R - contains script to modify original tracks related to metric A
- modify_metricB_tracks.R - contains script to modify original tracks related to metric B
- create_protocols.R - generates protocol files needed for the experiment
- prepare_results.R - will merge all response data collected from the experiment and generates plots

A.1.2 Data

Data folder contains two files,

- trackdata_mot_experiments.csv - contains trajectory data presented in previous MOT experiments
- all_responsedata.csv - contains response data collected from previous MOT experiments

A.1.3 Images and Videos

Folders named Images contains various plots generated during the project. Folder named Videos contains trajectory videos of the modified tracks.

A.2 Experiment

Folder named experiment contains file that runs the experiment using PsychoPy, data collected from the experiment, Motbox package that is necessary to run the experiment, protocol files, both modified and original trajectory files.