Learning to Catch:
Modeling Interception Success from Optic Cues

MASTER FINAL THESIS

MSc in Research in Behavior and Cognition

Universitat de Barcelona

Supervisor: Prof. Joan López-Moliner, PhD.

July 3rd, 2025

Robert Cuello López

NIUB 20368924

STATEMENT OF CONTRIBUTION TFM

Conceptualization: Generating ideas, formulating, or evolving overarching research goals and aims. □ **Joan López-Moliner, Robert Cuello López.**

Methodology: Developing or designing the methodology, creating models. □ **Joan López-Moliner, Robert Cuello López.**

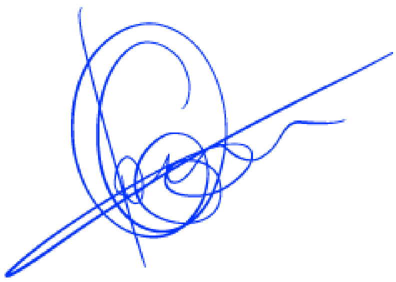
Software: Programming, software development, designing computer programs, implementing the computer code and supporting algorithms, testing existing code components. □ **Robert Cuello López.**

Investigation: Conducting research and investigation, specifically performing experiments or collecting data/evidence. □ **Borja Aguado, Joan López-Moliner**.

Formal Analysis: Applying statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data. □ **Borja Aguado, Joan López-Moliner, Robert Cuello López.**

Visualization: Preparing, creating, and/or presenting the published work, specifically visualizing and presenting data. □ **Robert Cuello López.**

Writing - Original Draft Preparation: Creating and/or presenting the published work, specifically writing the initial draft. □ **Joan López-Moliner, Robert Cuello López.**



Abstract

Interception tasks are a window to study the continuous interaction between perception and action. Models such as the Optical Acceleration Cancellation (OAC), Generalized OAC (GOAC), and Linear Optical Trajectory (LOT) have tried to explain how the interception movement is guided, arguing it is a real-time, constantly adjusted, heuristic-based control guided by optic variables. More recently, predictive models, such as the one proposed by Aguado & López-Moliner, suggest that interception behavior is guided through optic variables and internal priors, such as gravity.

This thesis aims to check whether this optic variables (angular size (θ) and elevation angle (pitch)) used to guide interception, contain outcome-relevant information, and if it is learnable by an artificial neural network (ANN). Specifically, the aims of the thesis are to: (1) check whether interception success can be predicted from optic variables, (2) how much input is required to retain predictive accuracy, and (3) whether gravity influences performance. Results show that the ANN achieved a moderately high predictive performance with high recall. Truncation analyses revealed that performance remains stable down to 35% of the input sequence, while removal of explicit gravity information slightly decreased accuracy. Notably, performance remained consistent across gravity conditions.

These results indicate that optic variables contain structured, outcome-relevant information, that are learnable by the ANN. Altogether, this results bridge heuristic models of control with predictive models, suggesting that optic variables can both be used to guide the interception behavior and to infer interception success.

KEYWORDS: Interception, optic variables, artificial neural networks, virtual reality.

INDEX

1. INTRODUCTION

Intercepting moving objects in real time is a deceptively simple behavior that remains a central challenge in behavioral neuroscience. Tasks such as catching a ball or swatting an insect require the brain to continuously integrate and act upon dynamic sensory input, often under a certain level of uncertainty or missing information. These behaviors require the coordination of perceptual processing, temporal estimation, and a timely motor execution.

Interception tasks are a valid and measurable model system for investigating the continuous interaction between perception and action. They involve sensorimotor integration, predictive processing, and the perception-action loop (Fajen & Warren, 2007). In other words, they allow researchers to observe the real time transformation of visual input into movement commands, illustrating the result of how the brain processes spatial and temporal cues to anticipate future states. A distinction needs to be made from static decision-making tasks, since success does not only depend on what is perceived, but rather on how it evolves over time.

A paradigmatic case of this research line is the ‘outfielder problem’: how does a baseball player —the defensive player, positioned the farthest from the batter, responsible to catch the high, long-range balls— manage to get to the right place on time and catch the flying ball? This underlies a broader question of how the brain solves spatiotemporal estimation tasks, especially given the limited and inherent uncertainty of sensory data at early stages of the fly. Furthermore, cues such as binocular disparity (the difference in images seen by each eye), texture gradient (changes in texture across a scene), or motion parallax (relative motion of objects at different distances) are minimized due to distance and / or speed, hampering the interception task (Cutting & Vishton, 1995). As a consequence, the brain must rely on other cues or models to guide action. Most importantly, the outfielder must solve where and when to act, integrating the temporal and spatial cues to guide the movements toward interception (Fink, Foo, & Warren, 2009).

To further explain how this estimation takes place in the brain, various models have been proposed to explain the basis of interception. These models can be divided into two main approaches: The first one, known as Trajectory Prediction (TP), describes how the observer perceives the initial characteristics of the movement —such as speed, angle, or acceleration— and with this information is able to predict the rest of the trajectory using internal models of projectile motion (Saxberg, 1987). This implies that the observer is capable of computing the influence of forces such as gravity or air resistance to estimate the exact endpoint. Although this might sound intuitively plausible, there is limited evidence supporting this model. Even highly trained athletes often fail to accurately estimate trajectories in novel or perturbed contexts, suggesting that such mental simulations may not be the primary control mechanism (Shaffer & McBeath, 2005).

In contrast, we find online control theories, which posit that interception estimations are achieved by the continuous coupling of motor output to optical variables obtained from the visual scene. Optical variables refer to the dynamic information that is already present on the retinal images. An example of these online theories is the Optical Acceleration Cancellation (OAC) model (McLeod & Dienes, 1996), which argues that the outfielder locomotion is regulated by canceling the tangent of the optical elevation angle through adjustments in the observers speed in order to maintain a stable value of the control value and consequently reach the interception point on time. It primarily explains the control of the object's in-depth movement. Furthermore, the generalized optical acceleration cancellation (GOAC) (McLeod, Reed, & Dienes, 2006) model expands and proposes maintaining a constant bearing angle in order to control the ball's horizontal motion. These two models do not assume any use of variables other than the optic ones, but rather rely on the perceptual system's sensitivity to changes in visual angles over time and their acceleration. On a similar page, we find the Linear Optical Trajectory (LOT) (McBeath, Shaffer, & Kaiser, 1995), which explains catching behavior by moving in a way to keep the object's image moving in a straight line on the retina. What these models have in common is that they suggest that effective interception is the result of real-time control of movement through the optically available signals, in contrast to the TP model, which is an inferential trajectory computation.

Moreover, these theories diverge in their predictions regarding behavior under perturbed or non-Newtonian trajectories. With the help of virtual reality (VR) studies, it has been possible to evaluate these differences by altering an object's dynamics mid-flight. For instance, Fink et al. (2009) implemented a VR setting that introduced mid-air changes in the vertical acceleration, and concluded that participants adjusted their movements according to the updated vertical acceleration rather than maintaining a precomputed path based on initial cues. These results were more consistent with the predictions made by the OAC and GOAC rather than the TP or LOT frameworks. This is compelling support in favor of the closed-loop perceptual control.

Together, the OAC, GOAC and LOT models emphasize how interception emerges as a continuous process of perceptual constraint satisfaction rather than a discrete decision-making event. Despite the difference between the models' approaches, they highlight the richness of information to be extracted by our optical senses regarding signals from angular motion cues.

Nevertheless, whilst these models objective is to account for and explain how the movement is guided toward the interception through optical cues, they do not assess whether the available perceptual information at any point throughout the object trajectory suffices to predict interception success or failure. This limitation grows larger, especially in scenarios where visual information is delayed, impaired, or suppressed. From here, a question is raised: How much of the optic variable sequence is needed before the interception outcome can be accurately classified? This question opens the door for testing how a predictive component present in the informational structure of early perceptual input can help solve the task.

Recent work by Aguado and López-Moliner (2025) began to bridge this gap by introducing a new perspective to the topic of interception. His study implemented a VR task involving ball trajectories which had their trajectories manipulated under certain conditions, such as gravitational acceleration and ball sizes. This resulted in a setting where they were able to have total control of the sensory environment whilst preserving ecological validity. Furthermore, their analysis went beyond traditional behavioral analysis; they implemented heuristics strategies, which use only optic variables to guide action, not requiring an internal model of physics or gravity. These strategies correspond to the models presented above, like the OAC and GOAC. In addition, Aguado et al. included a predictive/Bayesian controller, which assumes an internal model of how the ball moves (including gravity effect) and uses initial conditions (launch angle, velocity) to predict the ball's endpoint.

The findings revealed that interception behavior is sensitive to gravitational context. Furthermore, predictive models that incorporated a prior of gravity had a better fit to participant behavioral data than heuristic models had. On top of that, the study confirmed how specific optical variables are especially relevant in guiding interception. These are the angular elevation (γ, from now on pitch) and theta (θ) (see figure 1) —known as angular size or visual angle, which represents how large an object appears on the retina, depending on the size and distance of the observed object—. As a conclusion of the study, they suggested that humans do not only rely on perceptual coupling, but may also include latent knowledge about movement dynamics or gravity (Aguado & López-Moliner, 2025).

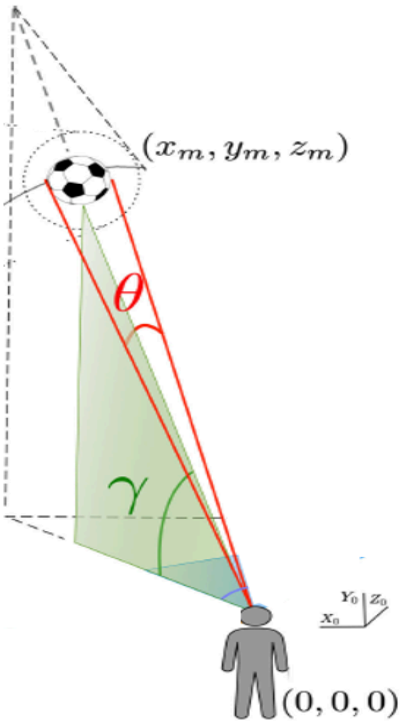


Figure 1: The figure represents the key optic variables: Elevation angle, also known as pitch (γ), and retinal object size (θ). Adapted from Aguado & López-Moliner (2025).

Nonetheless, the focus of Aguado & López-Moliner work was to reconstruct and compare movement strategies. They analyzed how participants moved in different gravitational situations and visual conditions. That is, focusing on execution rather than final outcome prediction. On a similar line of work, de la Malla & López-Moliner (2015) determined that integrating predictive and online visual information led to more accurate timing in interception, rather than using just one of the techniques.

The question of how much of the optic variable sequence is needed before the interception outcome can be accurately classified was not addressed. Answering this question would be a step toward understanding the informational value and temporal structure of optic variables during interception, providing a deeper understanding of the perceptual foundations of successful performance.

Building on the latest work of Aguado and López-Moliner (2025) and using their VR task dataset, this thesis aims to assess whether interception success can be predicted from early optic variables, before the full object trajectory has unfolded. While previous research has reached the conclusion that interception behavior adapts to gravitational acceleration and ball size through optical flow and internal physical priors, it still remains to check if an artificial neural network is able to learn and generalize this model of interception.

Thus, the central hypothesis is if early optic variables —reflecting both momentary visual input and prior knowledge about gravitational dynamics— contain sufficient information for the artificial neural network to accurately classify interception success. To further explore the dynamics of the encoding, the minimal input threshold hypothesis is also studied, where it will be assessed if interception success can still be predicted with a truncated segment of the optic variable sequence. In addition, it will be assessed whether classification performance varies across gravitational conditions.

2. Methods

2.1 Participants and task overview:

The data used in this study was obtained from the participants of an interception task in virtual reality, an experiment developed by Aguado & López-Moliner (2025). 12 participants took part (6 male, 6 female; around 22–23 years old). All participants had normal vision or corrected to normal vision, and did not know the goals of the study. Participant number 5 from the data analysis was excluded due to a technical issue in the head-mounted display (HMD), so the final sample consisted of 11 participants, and 9439 trials. For the classification task, the trials were randomly divided into 80% train and 20 % test sets, resulting in 1888 trials for the test set, from which the results are evaluated.

Participant role was to visually track and intercept balls that followed a parabolic trajectory under systematically varied conditions. The trials were divided into 4 phases:

1. **Launch** – The ball became visible and started its parabolic trajectory from a fixed distance of 40 meters across all trials.
2. **Flight** – The ball traveled toward its landing point.
3. **Occlusion** – In the final segment of the trajectory, the ball was not visible anymore, requiring the participants to complete the interception without the sight of the ball.

4. **Response** – The participants had to report their estimated time-to-contact by pushing a button when they believed the ball had reached again eye-level.

The systematically varied conditions of the balls parabolic trajectory were a product of three factors: gravity modification (8.8263g, 9.807g and 10.7877g, corresponding to the Moon, Earth, and Mars, respectively), physical size (0.198, 0.22, the equivalent to a soccer ball, and 0.242), and the landing point (6 different trajectory angles).

The data was recorded throughout all the trials at 90Hz using the integrated eye-tracker and position-tracking sensors of the HTC Vive Pro HMD, which yielded precise estimates of gaze direction and 3D position of both the participant and the ball at all times.

2.2 Data processing and feature construction

Only data from phase 2 (ball flight) was used for the model training. For each trial, the following variables were extracted.

- **G**: gravitational acceleration.
- **Size**: physical size of the ball.
- **pitchBall**: vertical angle from the observer to the ball (retinal elevation).
- **Theta, θ**: angle size of the ball on the retina, computed as $2 * \arctan\ arctan\left(\frac{radius}{distance}\right)$.

The data from each trial consisted of a sequence of θ and pitchBall values, with varying length depending on the flight duration. To allow batch processing in a neural network, a requirement is that all the input has the same shape. Therefore, all trials were standardized to the longest trial (622 time steps). Shorter trials were zero-padded to the length of the longest trial, which is standard practice when dealing with varying length inputs. Each trial was encoded as a feature vector:

$$X = [G,\ size,\ \theta 1, \ldots.,\ \theta 622,\ pitchBall1,\ \ldots,\ pitchBall622$$

The target label was binary, corresponding to whether the participant was successful at intercepting the ball (1) or not (0). Successful interceptions were calculated as the minimum Euclidean distance between the participant and the ball in phase 3. If the distance was equal to or smaller than 0.5 meters, it was considered a successful interception, otherwise, it was labeled as a miss.

The label definition prompted an imbalanced dataset, with approximately 30% successful trials and 70% misses.

2.3 Classification machine learning model

The classification model used to predict the outcome was a fully connected feedforward neural network implemented with PyTorch. Its architecture consisted of three hidden layers with the following characteristics.

- **Hidden layer sizes**: [128, 64, 64]
- **Activation**: Parametric ReLU (PReLU)
- **Batch normalization** applied after each layer
- **Dropout regularization**: p = 0.2
- **Loss function**: Weighted Binary Cross Entropy with logits (BCEWithLogitsLoss), using pos_weight = 2.33 to compensate for class imbalance
- **Optimizer**: Adam, with learning rate = 0.001 and weight decay = 1e-4
- **Batch size**: 64
- **Number of epochs:** 150
- **Classification threshold**: 0.6

2.4 Experimental design

To address the hypothesis of the thesis, the following experimental design was implemented:

H1. Full model performance: The classification model was trained using the full length of the input vectors (data corresponding to phase 2) and tested to held-out data.

H1a. Minimal input duration: the input vectors were truncated to the following lengths (75%, 50%, 40%, 35%, 30%, 20%, 10%) and zero-padded to match the original 622-length format. Equivalents to the previously described model were trained for each input length. To assess whether the models' performance (prediction quality) significantly decreased with the reduced input, bootstrap resampling (1,000 iterations) confidence intervals were used to compare F1 scores between the full model and the truncated models. In addition, McNemar's test was performed to evaluate whether 2 models (i.e.: 100% length vs. 35% length) made statistically different classification decisions on the same test set, independently of the F1 scores.

H1b. Gravity variation: 2 complementary approaches were followed:

1. The trained model was evaluated separately on test data grouped by gravity level. Then, performance metrics were obtained, and a chi-square test was used to compare prediction accuracy.
2. The model was retrained with the gravity input feature set to 0, and its F1 score performance was compared against the original model via bootstrap resampling (1,000 iterations) confidence intervals.

2.5 Evaluation metrics:

The following metrics were computed on the test sets and used to evaluate and compare model performance. :

- **F1 Score**: Harmonic mean of precision and recall, especially informative for imbalanced classification.
- **AUC (Area Under the ROC Curve):** Measures the model's discrimination ability across all thresholds.
- **Binary Cross Entropy Loss**: Optimization objective during training.
- **Precision-Recall Curve:** Visual evaluation of the trade-off between precision and recall across different thresholds.
- **Recall (Sensitivity):** Proportion of actual catches identified by the model. In terms of signal detection theory (SDT), this corresponds to the hit rate, calculated as: *True positives / (True positives + False negatives).*
- **Precision**: Proportion of trials predicted as catches that were correct, that is, the reliability of the model when it predicts a catch. In SDT terms, it is related to how well the model avoids False Alarms -predicting a catch when it was not, a false positive-. Can be interpreted as a False Alarm control, since the higher the precision, the fewer incorrect catch predictions (false alarms). It is calculated as: *True Positives / (True Positives + False Positives).*

To further contextualize precision and recall into the signal detection theory, let's see the following example:

Imagine in a test there are 1000 trials, of whom 300 are true signals (i.e: catches), and the model predicted 400 signals.

Out of these, 240 were correct, i.e., hits. Another 60 signals predicted were noise, i.e., false alarms. Therefore, 60 signals were missed, i.e., misses. Finally, 700 – 60 = 640 noises were correctly rejected, i.e., correct rejections.

So, *Recall = Hits / Hits + Misses*, how sensitive is the model at detecting real hits, hit rate.

On the other hand, *Precision = Hits / Hits + False Alarms*, how reliably it predicts Hits, or how much it minimizes false alarms.

2.6 Implementation details:

All models were implemented in PyTorch, and run on a laptop standard CPU (Intel i5). While the code was designed to be device-agnostic and is able to leverage GPU acceleration if available, no GPU was used. To ensure reproducibility, 'torch.manual_seed(42)' was fixed at the start of each training run. All computations were done in Python 3.12.7, within Jupyter Notebook scripts, following a consistent data pipeline. Visual Studio Code was used as the text editor. Regarding the data extraction and preprocessing from the Aguado & López-Moliner (2025) dataset, it was done using R, and RStudio as the IDE.

3. Results

This section focuses on the performance of the classification machine learning models developed to predict interception based on optic variables. The reported results are the outcomes of the three hypotheses tested: (H1) predictability of interception outcome using only optical variables, gravity, and size, (H1a) the effect of reduced optical input duration, and (H1b) the influence of gravity on the model performance.

3.1 Predictive performance of the interception outcome

To evaluate whether interception outcomes are predictable with optic variables and gravity, a fully connected forward neural network was trained using the full length of optic variables from phase 2. The model steadily improved throughout training. At the final epoch, the model achieved the following results on the held-out test data:

- **F1 Score**: 0.587 (figure 2, panel C)
- **Area Under the ROC Curve (AUC)**: 0.747 (figure 2, panel D)
- **Recall**: 0.783 (figure 2, panel E)
- **Precision**: 0.470 (figure 2, panel F)
- **Binary Cross Entropy Loss**: 0.9124 (figure 2, panel B)

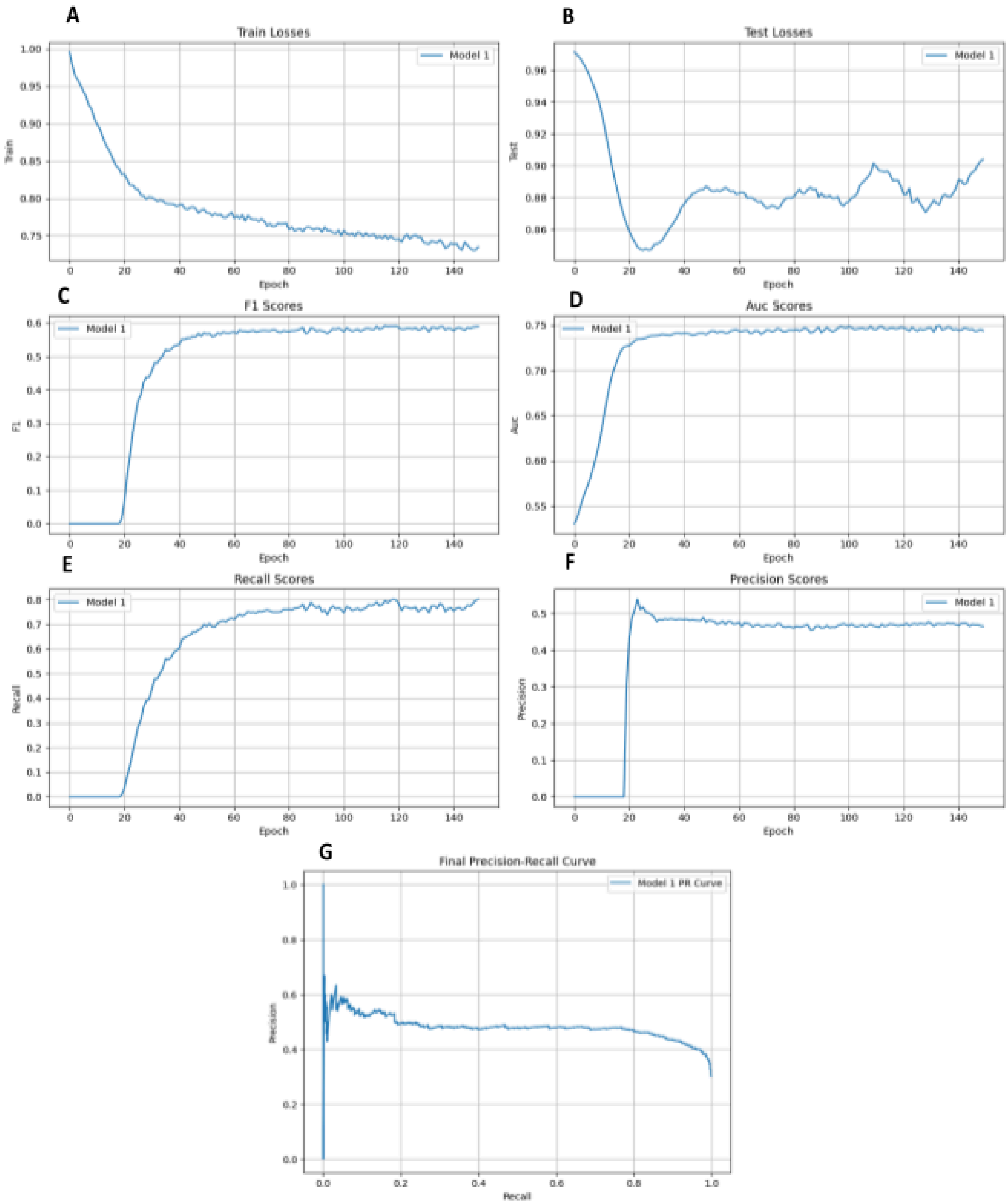


Figure 2: *Main model performance metrics. A and B are test and train losses respectively, which shows model convergence and fit to data. C is the F1 score, which reflects the performance of the model. D, is the Area Under the ROC Curve, informs about how good the model is at distinguishing between the two classes. E is the proportion of correctly identified instances of the positive class (i.e., catches), the hit rate. F is precision, the reliability of the model when predicting an instance of the positive class (i.e., catches). G is the precision-recall curve, which provides a view of the model's performance at different thresholds.*

Table 1: Confusion matrix of main model predictions

The confusion matrix at the last epoch was:	Predicted: Miss	Predicted: Catch
Actual: Miss	813	504
Actual: Catch	124	447

The results (see figure 2) show how, in the test trials, the model succeeded in capturing significant predictive information from the optical variables. The high recall (0.783) (see figure 2, panel E) demonstrates that the model is able to correctly identify a large proportion of successful interceptions. On the other hand, the precision score at 0.470 (see figure 2, panel F) shows a higher false positive rate, which is to be expected given the imbalanced data and the higher cost of missing a catch. The F1 score achieved was 0.587 (see figure 2, panel C), indicating balanced performance in detecting true positives while minimizing false positives and false negatives. The AUC score of 0.747 (see figure 2, panel D) shows a moderately high separability between successful and not successful trials. Finally, the loss at 0.9124 (see figure 2, panel B) reflects convergence and moderate predictive certainty in probability estimates.

3.2 Effect of reduced input length (H1a)

In order to evaluate how much of the optic input length is necessary for reliable prediction with a performance statistically similar to the one of the main model, the model was trained using a truncated version of the input, which corresponded to 75%, 50%, 40%, 35%, 30%, 20%, and 10% of the original input.

To assess if the model performance was retained at the shorter input lengths, bootstrap confidence intervals (CI) were used. This tool was applied to the F1 score differences between the full length model and its truncated versions. The following results were obtained:

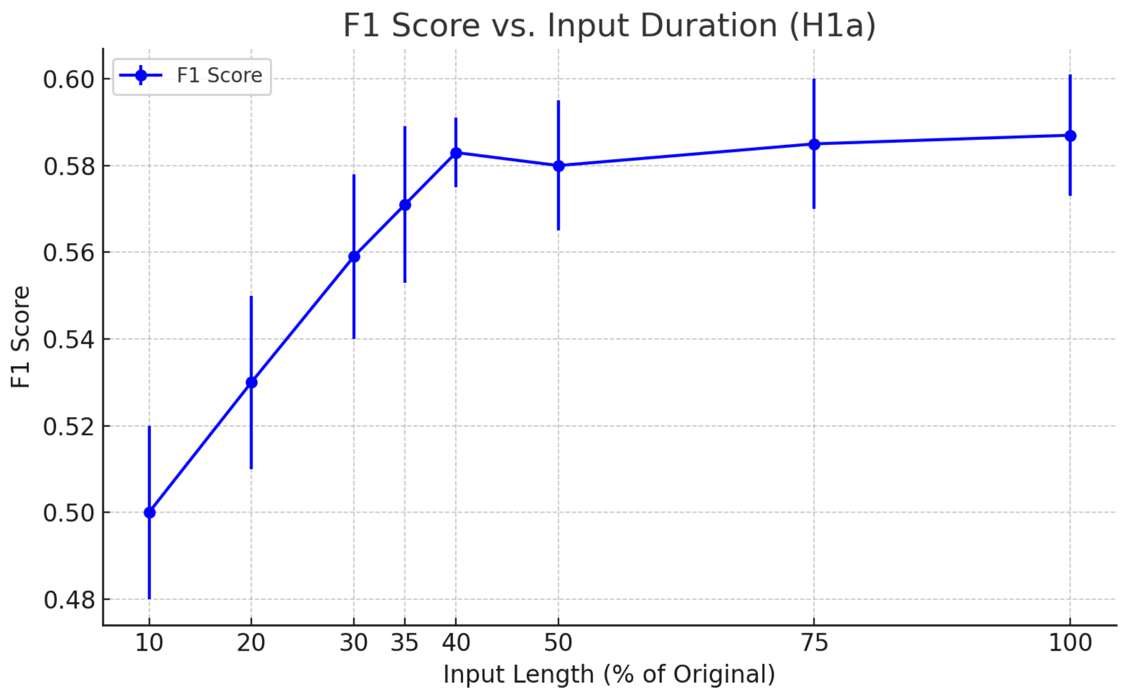


Figure 3: *F1 scores by truncation length*

After visual inspection of figure 3, there is an evident drop in F1 performance at 35%, and even more accentuated at 30%, and so on. To statistically confirm its significance, the obtained CI were evaluated, as shown in the table below.

Table 2: 95% CI of the F1 score between the truncated models vs. full length model

Truncation Level	95% CI for F1 Difference vs. 100%	Significant?
35%	[-0.0084, 0.0319]	**No**
30%	[0.0173, 0.0490]	**Yes**

Note. Significance is assessed by the confidence interval overlapping 0 or not.

At 35%, the CI does include 0, hence the difference between models is not statistically significant. On the other hand, the 30% CI does not include 0, hence, this truncated model performance is deemed significantly different from the full length model. Nevertheless, the narrow CI suggests that the statistical difference is not large.

In addition, McNemar's test was performed to indicate whether the truncate inputs led to different classification decisions.

Table 3: McNemar's test comparing classification decisions between full length model and truncated models

Model Comparison	McNemar p-value	Significant Difference in Decisions?
100% vs. 35%	.650	No
100% vs. 30%	.313	No

Note. Significance based on $\alpha = .05$.

Although model performance significantly dropped at the 30% truncated length, the classification decisions remained statistically consistent, suggesting a drop in performance rather than a decision shift.

3.3 Role of gravity in predictive performance (H1b)

The final hypothesis examined whether gravity influenced model performance. To address this final hypothesis, three approaches were implemented: evaluation of model performance under different gravities, comparison of prediction correctness across gravity levels, and model performance evaluation and comparison without gravity input.

3.3.1 Performance by gravity level

Unique gravity values were segmented (G = 8.8263, 9.807, and 10.7877), and model performance was evaluated for each. Results are shown in figure 4:

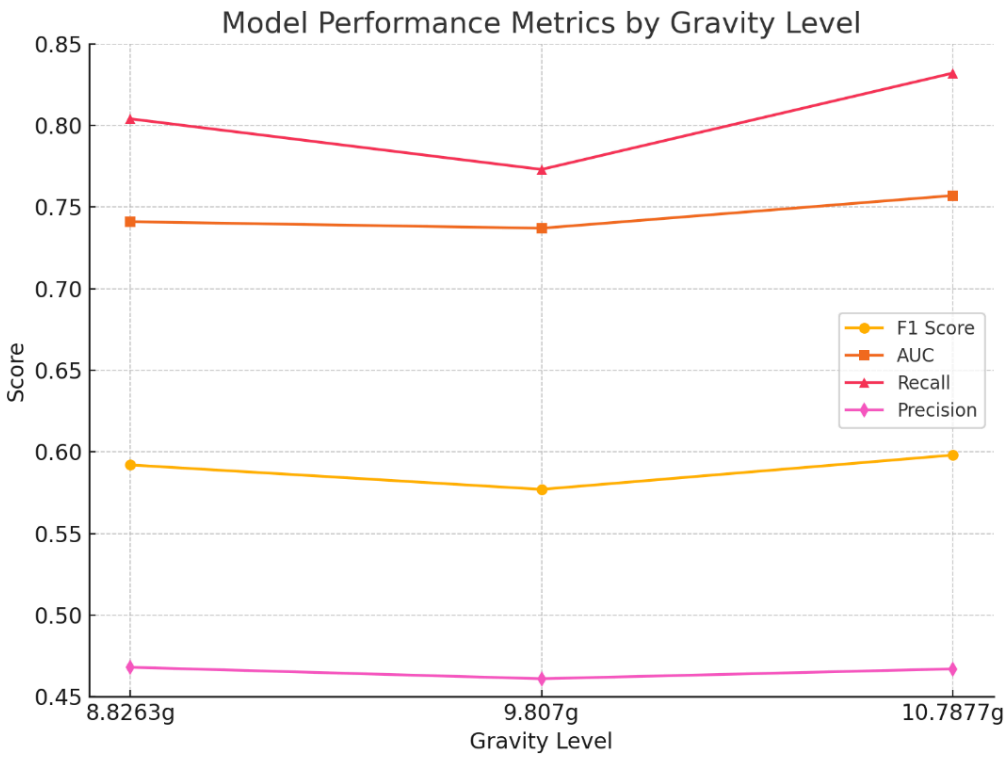


Figure 4: *Model performance metrics under different gravities.*

The graph alone does not show a substantial difference in model performance across gravity conditions, but, to formally evaluate this statistical significance, a chi-squared test of independence was performed on the prediction correctness across gravity levels, using the following contingency table:

Table 4: Model correct and incorrect predictions of interception success under different gravities

Outcome	*G = 8.8263*	*G = 9.807*	*G =10.7877*
Correct	**422**	**408**	**419**
Incorrect	**215**	**224**	**200**

This contingency table yielded the following chi-squared results: $\chi^2(2) = 1.38$, $p = .503$. Thus, there is no significant association between gravity and prediction performance.

3.3.2 Gravity vs. No-gravity models

To assess whether gravity has a significant contribution to model performance, the model was retrained with the gravity input set to 0 to eliminate any gravitational information. The final test performance was:

- F1 Score: 0.557
- AUC: 0.745
- Recall: 0.701
- Precision: 0.462

Comparison between this model and the model with gravitational information performance in terms of F1 score is shown in figure 5:

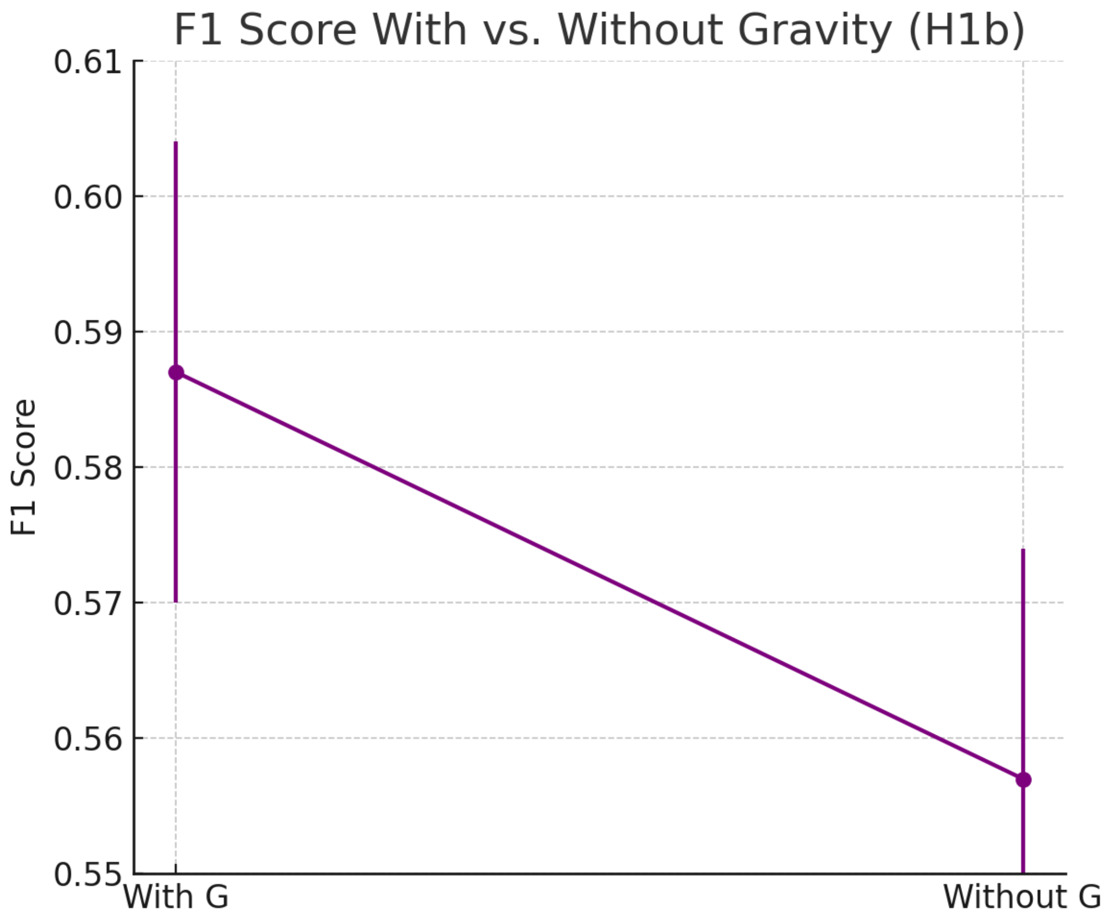


Figure 5: F1 score comparison between gravity vs. non-gravity models

Again, to statistically endorse whether there is a significant difference between the models' performance, a bootstrap CI was performed for the F1 difference:

95% CI for F1 difference (with G vs. without G): [0.0173, 0.0490]

Thus, we can confirm that, by means of the bootstrap CI not including 0, there is a significant difference in terms of F1 scores between the models.

4. DISCUSSION

The goal of this thesis was to investigate whether early optical variables –concretely pitch and visual angle theta (θ)– could predict interception success in a VR task, and whether this predictive ability was dependent on the length of the input available or the gravitational conditions. This approach builds on how interception movements are executed, and shifts toward what information is available early enough to predict the likely outcome. Results showed that a supervised neural network trained on optical cues is able to predict interception success with moderate performance. Moreover, it showed how, even with the input truncated, it was still able to make correct classifications. Although gravitational input proved to have a slight enhancement on performance, it was not essential for accurate predictions, and no significant differences between the studied gravities were observed. These findings suggest that there is outcome-related information available encoded in the optic variables, which, when coupled with the internal prior, jointly contribute to successful interception, in line with the findings of de la Malla & López-Moliner (2015), who showed that humans rely on both optic variables and internal priors, such as gravity models, to aid in interception.

Optic variables encode outcome-related information learnable by an artificial neural network

The main finding of this work is that the artificial neural network is able to learn how optic variables alone are enough to predict whether interception will succeed. Getting into the specifics, when analyzing the unseen trials, that is, the test trials, the network achieved a reasonably high recall (0.783) and moderate AUC (0.747). These performance metrics argue in favor of how dynamic perceptual information, angular elevation, and visual angle (θ), encode relevant structure regarding the outcome of the interception. The proven ability of the model to generalize to unseen data indicates how the model was able to learn the embedded patterns in the data rather than just memorizing specific trajectories.

These findings are aligned with the broader perspective in the perceptual-motor literature that posits that optic flow contains rich and usable information to guide behavior (Fajen & Warren, 2007). Interception models such as the OAC and the GOAC (McLeod & Dienes, 1996; McLeod et al., 2006) describe how movement is continuously modulated by the optic variables. The findings of this work argue that these same variables, also encode enough relevant information regarding the interception outcome, and hence it can be extrapolated that the perceptual input of the online control also contains sufficient information to infer interception success.

It is important to note that the network's predictive success implies that angular measures, when sampled over time, do reflect multiple latent factors, such as approach speed, time-to-contact, or early divergence in motor coordination. As a consequence, although the network was trained without explicit access to the internal motor state or the intentions the participants bore, the optic variables alone were informative enough to yield an above-chance performance.

Temporal truncation reveals the minimum input required for predictive performance.

Another goal of the study was to shed light on whether there exists a minimum length required for interception successful prediction without statistically losing performance to the full length input. To test this, truncated sequences of the input were fed to additional models, and F1 scores were statistically compared to the full input model. Results showed how, up until 35% of the trajectory duration, the predictive performance was not significantly reduced. However, at 30% of trajectory duration, the predictive performance dropped significantly, although the difference was not large.

Thus, the observed statistical threshold around 30-35% mark indicates that there exists a key mass of perceptual information needed to disentangle the likely outcome. This temporal mark may reflect the necessary time for meaningful divergence between successful and unsuccessful interception attempts to be encoded in the optic signal.

Interestingly, despite the significant difference in F1 scores, McNemar's test indicates how, in both the full and truncated models, the classification criteria remained the same, and only overall performance decreased. This supports the idea that core discriminative characteristics are already present in the foremost early optic variables, and longer inputs serve the objective of refining rather than defining the predictive basis. The evidence is congruent with findings in the perceptual decision-making realm, where early sensory evidence can report robust initial estimates, which will be later refined by the following input (Gold & Shadlen, 2007).

Gravity enhances prediction accuracy rather than determining it.

Another addressed question was the role of gravity in the predictive performance. Results showed how the model without any explicit gravitational information had a small but significant drop in F1 score performance, although the AUC and recall scores remained relatively stable. Furthermore, no significant differences in model performance across the three gravity levels (8.8263, 9.807, and 10.7877) were observed, just like no alteration in the proportion of correct predictions was observed in a chi-squared analysis.

These results suggest that although gravity contributes to model performance –most likely by aiding in structuring the internal angular changes over time– it is not essential for making accurate classifications. This is due to the optic variables already reflecting the gravitational effects indirectly through their temporal evolution. Thus, the inclusion of gravity seems to refine rather than enable successful classification.

The findings stay in line with the work of Aguado & López-Moliner, who showed that behavior is adapted based on gravity, and internal models that incorporated gravitational priors better accounted for observed trajectories. Nevertheless, while their work focused on movement execution, the present work focuses on information sufficiency, that is, whether the perceptual input contains implicitly enough information about the outcome, so an artificial neural network is able to learn it. The presented evidence suggests information sufficiency, regardless of gravity explicitly provided.

Implications of the findings

The results found further reinforce how interception is not solely a process of continuous action control, but how its early perceptual information carries predictive value. This supposes a conceptual bridge between models that focus on optical control –i.e., OAC, GOAC, LOT– and models which propose internal predictive components.

The fact that the artificial neural network has been able to successfully learn outcome-relevant patterns from the optic inputs indicates how this information is embedded in the optic variables, and its structure is detectable and learnable by a non-biologic learner. While this evidence is far from implying nor explaining that the human brain uses the same strategy, it still provides indirect evidence that such structure exists and is learnable.

Limitations and future directions

Despite the interesting results, there are a number of limitations that we must draw attention to. Firstly, the data was obtained from a task in a virtual reality settings, which allows for great experimental control, but may not capture the complexity and noise from a real environment. This supposes a constraint when trying to generalize to real-world interception tasks.

The second detected limitation lies in the nature of artificial neural networks. While it revealed that prediction is possible from optic variables, it does not explain how those predictions are achieved, and neither does it explain how the information is dealt with or organized by the perceptual human system.

Thirdly, the present study did not test whether the learned model is able to generalize across individuals with different skill levels or strategies, hence limiting external validity.

To complement or expand the present work, the present artificial neural network model could be compared against biologically inspired models, such as recurrent or predictive coding networks, in order to try to bridge the gap between computational and cognitive plausibility. This approach may offer more insights on how humans process and act upon visual information in real time.

On top of that, testing the present model with perturbed or occluded trajectories would provide more solid conclusions about its predictive flexibility, and simulate better real-world uncertainty. Introducing these mid-flight changes would help distinguish whether the model relies more on momentary information or sequential temporal patterns.

Data accessibility*: The data, the R code for data preparation, and the Python scripts for model training and data analysis are available in the GitHub repository (Cuello, 2025).*

Acknowledgments: *I would like to thank my thesis supervisor, Prof. Joan López-Moliner, PhD, for his guidance and support throughout the thesis writing. Also, I would like to thank the Vision and Control of Action (VISCA) Group at the University of Barcelona for its feedback about the thesis.*

REFERENCES

Fajen BR, Warren WH. 2007 Behavioral dynamics of intercepting a moving target. Exp. Brain Res. **180**, 303–319. (doi:10.1007/s00221-007-0859-6)

Cutting J, Vishton P. 1995 Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. In Perception of space and motion, pp. 69–117. Cambridge, MA: Elsevier. (doi:10.1016/B978-012240530-3/50005-5)

Fink PW, Foo PS, Warren WH. 2009 Catching fly balls in virtual reality: a critical test of the outfielder problem. *J. Vis.* **9**, 14. (doi:10.1167/9.13.14)

Saxberg BVH. 1987 Projected free fall trajectories. *Biol. Cybern.* **56**, 159–175. (doi:10.1007/BF00317991)

Shaffer DM, McBeath MK. 2005 Naive beliefs in baseball: systematic distortion in perceived time of apex for fly balls. J. Exp. Psychol. **31**, 1492– 1501. (doi:10.1037/0278-7393.31.6.1492)

McLeod P, Reed N, Dienes Z. 2006 The generalized optic acceleration cancellation theory of catching. J. Exp. Psychol. **32**, 139–148. (doi:10.1037/0096-1523.32.1.139)

McLeod, P., and Dienes, Z. (1996). Do fielders know where to go to catch the ball or only how to get there? J. Exp. Psychol. Hum. Percept. Perform. **22**, 531–543. doi: 10.1037/0096-1523.22.3.531

McBeath MK, Shaffer DM, Kaiser MK. 1995 How baseball outfielders determine where to run to catch fly balls. Science **268**, 569–573. (doi:10.1126/science.7725104)

Fink PW, Foo PS, Warren WH. 2009 Catching fly balls in virtual reality: a critical test of the outfielder problem. J. Vis. **9**, 14. (doi:10.1167/9.13.14)

Aguado B, López-Moliner J. 2025 The predictive outfielder: a critical test across gravities. R. Soc. Open Sci. **12**: 241291. https://doi.org/10.1098/rsos.241291

de la Malla, C., & López-Moliner, J. (2015). Predictive plus online visual information optimizes temporal precision in interception. Journal of Experimental Psychology: Human Perception and Performance, 41(5), 1271–1280. https://doi.org/10.1037/xhp0000075

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. Annual review of neuroscience, 30, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Cuello, R. (2025). Learning to Catch: Modeling Interception Success from Optic Cues (v1.0) [Source code]. GitHub. https://github.com/Robert-Cuello/Learning-to-Catch-Modeling-Interception-Success-from-Optic-Cues.git