



HAL
open science

Analyzing Gaze Behaviors in Interactive VR Scenes

Kateryna Pirkovets, Clément Merveille, Florent Alain Sauveur Robert, Vivien Gagliano, Stephen Ramanoël, Auriane Gros, Hui-Yin Wu

► **To cite this version:**

Kateryna Pirkovets, Clément Merveille, Florent Alain Sauveur Robert, Vivien Gagliano, Stephen Ramanoël, et al.. Analyzing Gaze Behaviors in Interactive VR Scenes. NeuroMod 2023 - Annual Meeting of the NeuroMod Institute, Jun 2023, Antibes, France. hal-04208129

HAL Id: hal-04208129

<https://inria.hal.science/hal-04208129v1>

Submitted on 15 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



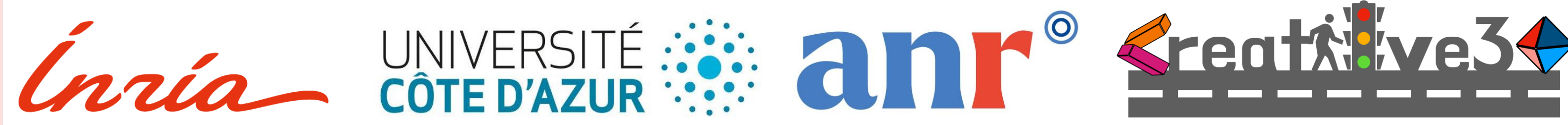
Distributed under a Creative Commons Attribution 4.0 International License

Analyzing Gaze Behaviors in Interactive VR Scenes

Kateryna Pirkovets¹, Clément Merveille¹, Florent Robert^{1,2}, Vivien Gagliano¹, Stephen Ramanoël^{3,4}, Auriane Gros⁵, Hui-Yin Wu¹

1 Centre Inria d'Université Côte d'Azur
 2 Université Côte d'Azur, CNRS, I3S
 3 Université Côte d'Azur, LAMHES
 4 Sorbonne Université, INSERM, CNRS, Institut de la Vision
 5 Université Côte d'Azur, CHU de Nice, CoBTek

Correspondence: hui-yin.wu@inria.fr



Highlights

- ❖ Understanding **visual attention in VR with free movement (6 degrees of freedom, free walking)**
- ❖ Varying stimuli: **2D and 3D, animated and static, interactive**
- ❖ Study with **12 participants in a 3.5m x 3.5m space**
- ❖ Implementation and evaluation of **six gaze behavior identification algorithms**
- ❖ Analysis using **linear mixed models** to investigate impact of stimulus, algorithm, and participant on gaze durations

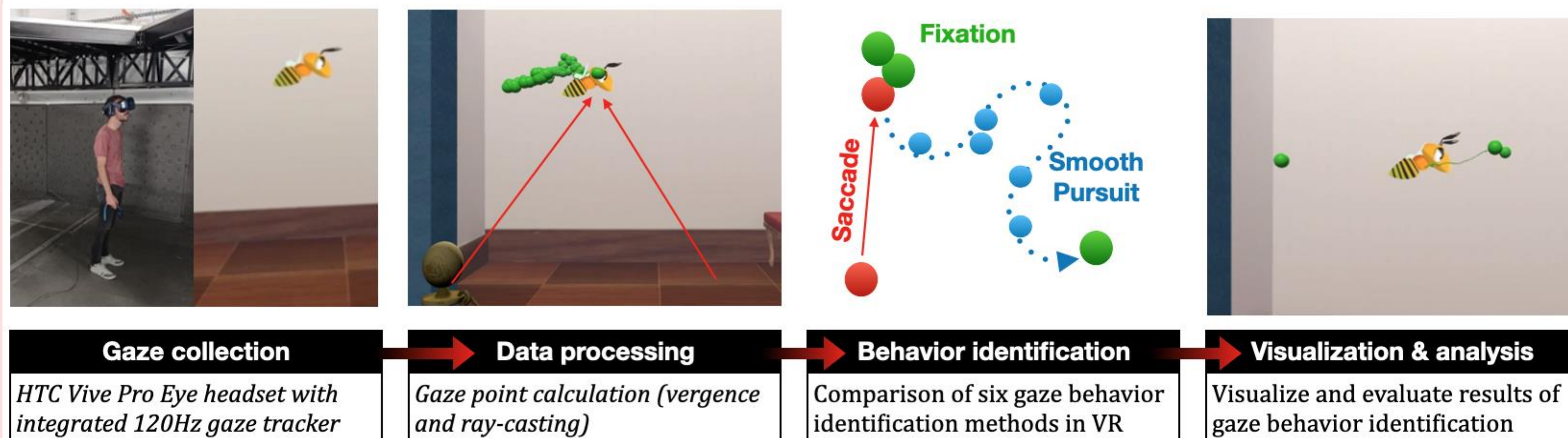
Summary

Gaze is an excellent metric for understanding human attention. However, research on identifying gaze behaviors – saccades, fixations, and smooth pursuits for example – for large (i.e., more than one meter viewing distance), interactive 3D scenes with virtual reality headsets is still in its early stages. The understanding of gaze behaviors is affected by equipment, user, the types of scenarios targeted, etc. There is currently little to no consensus on how to select gaze behavior identification methods, and what the impact of this choice has on the validation of research hypotheses on human attention.

This work investigates the impact of gaze behavior identification approaches on human gaze data analysis by re-implementing six state-of-the-art identification algorithms for VR. To underline the potential of the system, we design a 3D scene with various animated and interactive 2D and 3D stimuli with which we collected eye tracking data for 20 participants through a study. We then provide disaggregated analyses on metrics from the literature for various methods and stimuli.

From what we observe with a current state of analysis, participants tend to have longer fixations on dynamic 3D stimulus than on static ones, while some algorithms (based on their nature) fail to detect fixations on specific stimuli for some of the participants. We should also underline the differences in fixation patterns for participants as their fixations differ in quantity and duration across the algorithms.

Workflow



Our workflow consists of four steps:
 (1) **collection of gaze data** in an immersive VR scene,
 (2) **processing the data** to calculate the gaze points (which we refer to as Points of Regard, PoR) from gaze vectors,
 (3) **classifying gaze points** into various **gaze behaviors** for which we implement six different identification methods from literature, and finally
 (4) **visualizing results and analyzing metrics** for various gaze behavior identification methods.

Gaze behavior identification

Velocity threshold identification (I-VT) algorithms are based on the fact that the eye moves much more quickly during a saccade than during a fixation. In the existing implementations of I-VT, the velocity threshold has been set using angular degrees per second *angle/1t*

Variations implemented:

- **I-VT classic**: velocity thresholds tested include 30 (used by Tobii eye trackers; www.tobii.com), 120 (Salvucci & Goldberg, 2000) to 150 (Chen & Hou, 2022) angular degrees
- **REMoDNaV** (Dar et al., 2021; Liebers et al., 2021): A velocity type algorithm with adaptive thresholds and has been used for gaze data on 3D videos.
- **ISST** (Callahan-Flintoft et al., 2021): considers the velocity of both the head and the eyes. ISST classifies two types of gaze behaviors in addition to fixations, saccades and smooth pursuits: *head pursuit* and *vestibulo-ocular reflex (VOR)*.

Dispersion threshold identification (I-DT) Dispersion threshold identification (I-DT) assumes that eye movements during the fixation will be closely clustered. Most recent algorithms use the gaze vector angle, but the point position can be used when the head or eye position is not known.

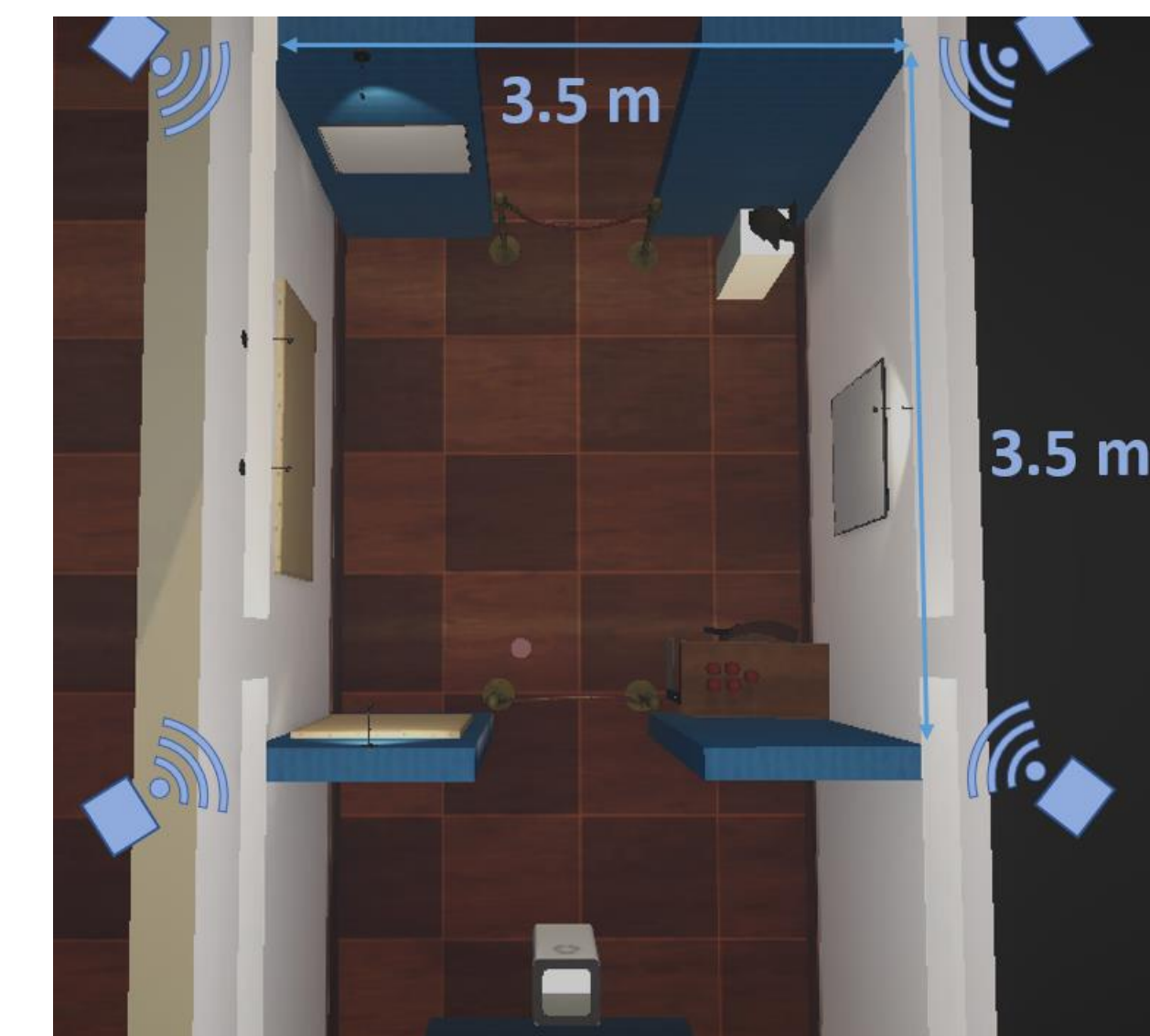
Variations implemented:

- **I-DT classic**: with a sliding window of 100-200 ms, two means clustering of gaze points per window, and StDev=0.5 for the threshold for identifying if a fixation is present
- **I-DTVR** (Llanes Juaros et al.2020): takes a time window and looks if the minimum frequency between samples in this window is higher than a threshold and if the maximum distance between two points is lower than a dispersion threshold. If yes, we consider all the points of this window as a fixation and extend the window until one of the conditions is no longer met.

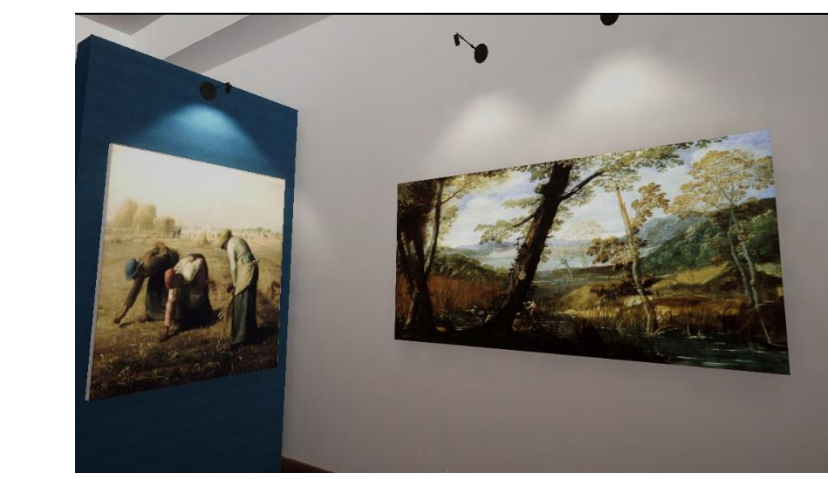
IVDT (Chen et al. 2022) A hybrid approach using both gaze velocity and clustering to identify fixations. The algorithm first classifies the sample point according to its velocity, then it checks if the fixation groups have the necessary duration and that each point has a distance to the other points lower than the dispersion threshold. If this is the case, the window is considered as a fixation.

Methods and stimuli

We implemented the entire workflow in Unity 2021.3. User gaze data is collected using the HTC Vive Pro Eye headset which has integrated gaze tracking and uses four base stations to track the headset position. The eye tracking data is collected through the SRanipal SDK at 120 Hz, and head motion is collected through callbacks in Unity at 90 Hz. The headset is tethered to a laptop equipped with a NVidia GTX3080 graphics card and an Intel Core I9 11950H CPU.



The experiment is setup with a 3.5 by 3.5-meter navigation space, with steam base stations at each corner of the space, as shown in Figure 4. The scene is designed in such a way that the users would not exceed the scene boundary to interact with different stimuli, though the virtual grid of Steam VR becomes visible if the user approaches it.



Two static 2D paintings: (Left) Jean-François Millet's *The Gleaners* with salient subject and (Right) Annibal Caracci's *Paysage avec Rivière* with no salient subject matter

2D animated video *Big Buck Bunny* (2008)



Low-poly 3D model of a rabbit



Abstract 2D video from Heider and Simmel (1944)

3D animated flying bee



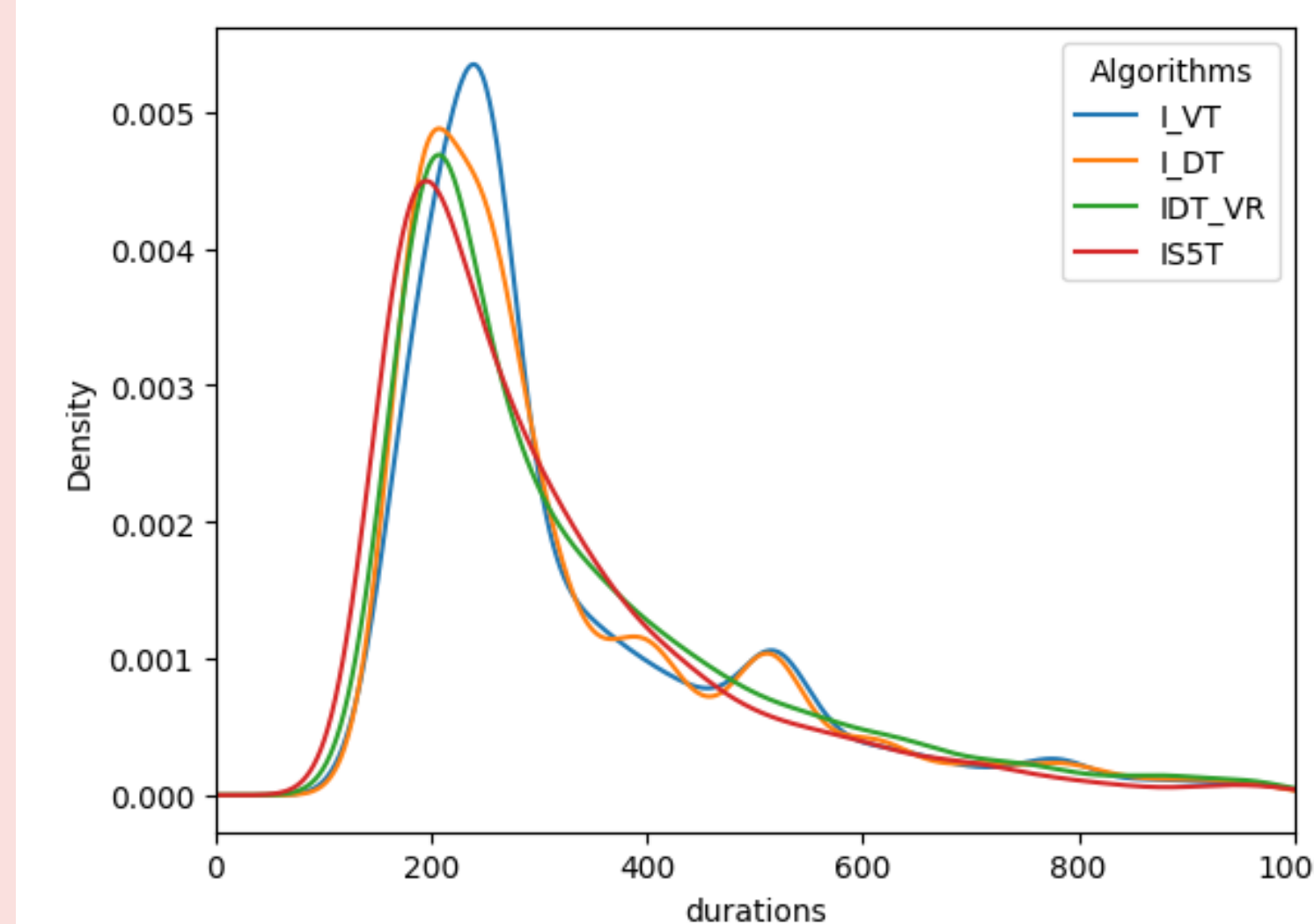
3D ball throwing game



Using **immersive virtual reality**, we observe **individualized gaze patterns** across participants in varied visual stimuli.

We also highlight through **statistical modeling** the impact of state of the art **gaze identification methods** on the analysis of gaze behavior.

Results

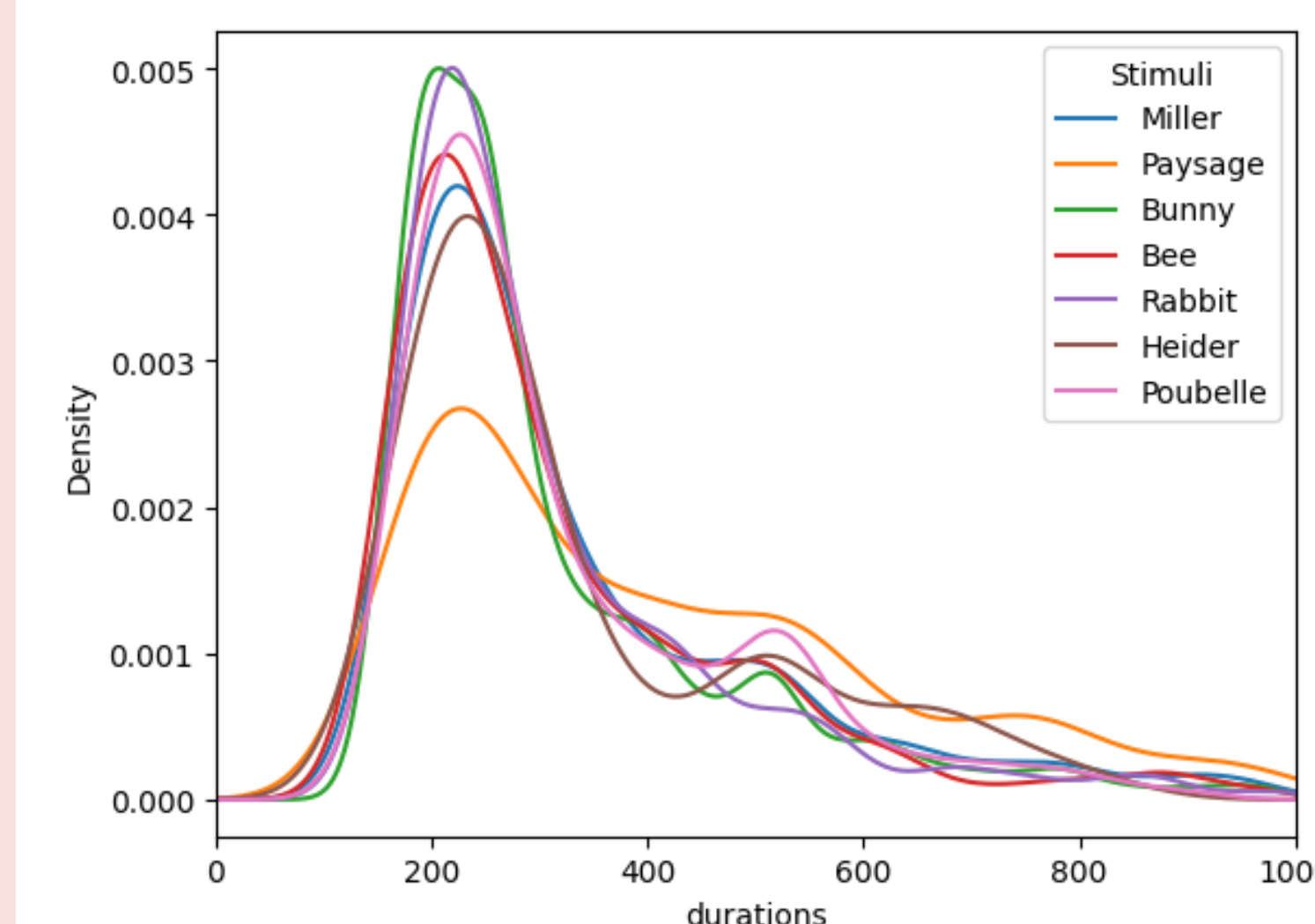


Most fixation durations have length ~250msec, algorithms show similar results.

Distribution is right-skewed.

Classical algorithms (I-VT, I-DT) have highest peak.

We keep the fixation between two-value (min and max) and the mean proportion of fixation discarded using this approach.

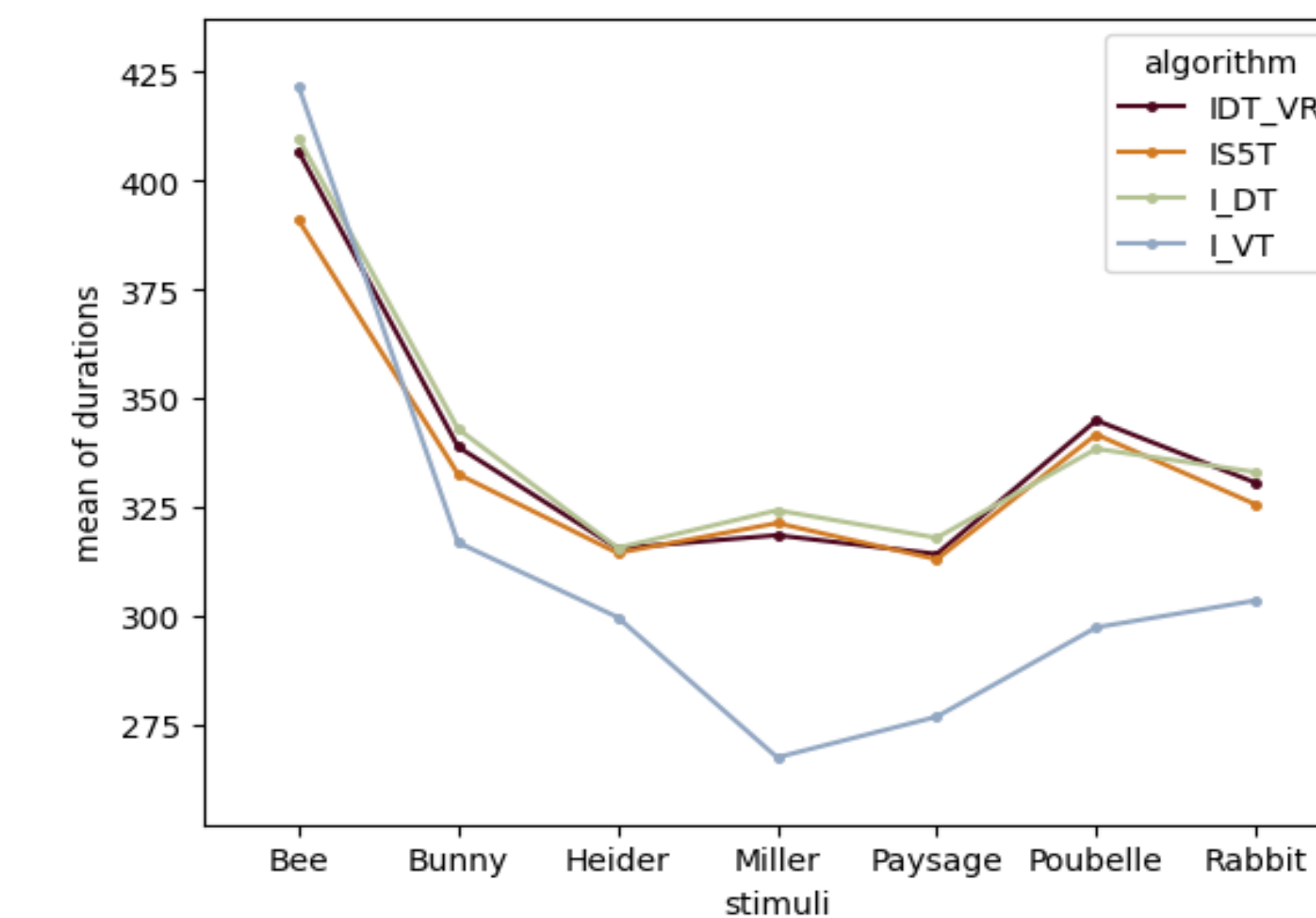


The upper limit at 1sec was set according to the literature, which removed less than 5% of all detected fixations for most of the stimuli.

People tend to have shorter fixations on 3D objects.

No visible difference in distributions between static and dynamic stimuli.

Distribution of fixation duration is right-skewed.

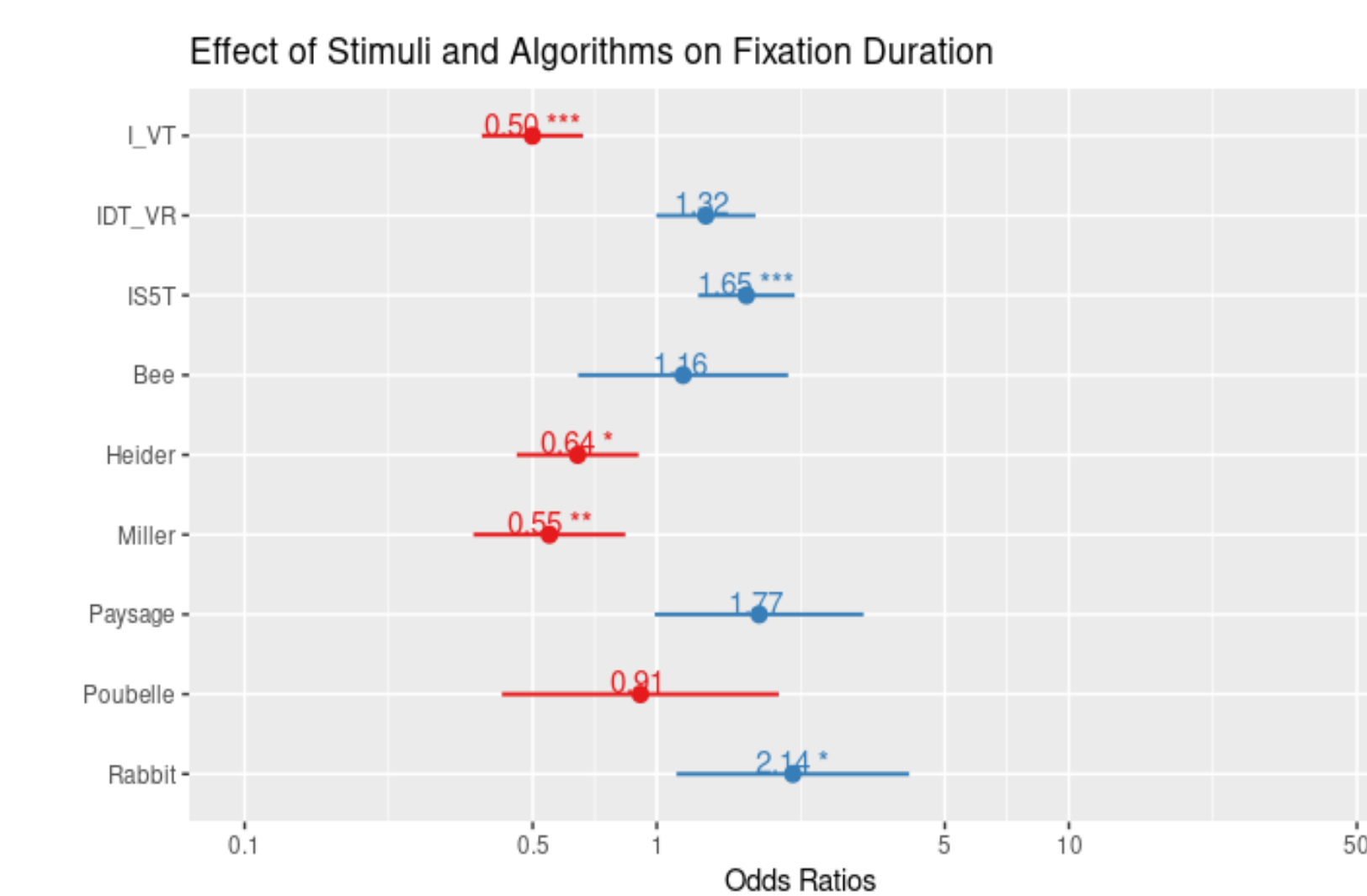


Algorithms are not consistent across stimuli.

All of them identify longest fixation durations for 3D moving stimulus (Bee).

All algorithms show similar results for each stimulus, except for I-VT.

We also didn't take into account short fixations (less than 30msec), and analyzed only those fixations, which were detected during the specific task.



Model: durations ~ algorithm + stimuli + (1 | participant)

We had to proceed with mixed models because data points cluster between individuals, therefore our data is not independent and can't satisfy linear models' assumption. This choice of model is the best one based on the model selection (AIC, BIC criterion).

Algorithms: I-VT, ISST (velocity based) and stimuli: Heider, Miller and Rabbit have significant impact on the fixation duration.