Andy Kapoor

2 May 2024

The purpose of this analysis was to assess temporal autocorrelation in gaze fluctuations across two mechanistically different virtual reality (VR) games. The study by Ainzenman et al. focuses on analyzing eye-tracking data gathered from players engaged in multiple VR games. In VR gaming, players' visual focus can significantly alter their performance based on game mechanics. Eye-tracking technology provides a link between game design and user engagement. The primary question this analysis seeks to answer is whether temporal autocorrelation patterns exist in the gaze fluctuations of a participant while playing different VR games, and whether these patterns vary across games. The null hypothesis posits no temporal autocorrelation patterns in the fluctuations of gaze direction across different VR games, suggesting that gaze movements are random over time within each game. The alternative hypothesis proposes existence of temporal autocorrelation patterns, suggesting that gaze movements are not random and may follow specific trends or cycles.

Gaze fluctuations were measured as immediate eye direction changes over time, to inform how players react to game stimuli in the VR environment. Autocorrelation is correlation of a signal in a delayed copy of itself. In the hypothesis, it implies that current gaze data points are correlated with past gaze data points. Identification of autocorrelation patterns would indicate presence of detectable and potentially predictable forms or trends in the data to suggest regularity or repeatable arrangements in gaze data across time. This analysis focuses on two rhythm-based games, Beat Saber and Pistol Whip, as these two games have contrasting interaction mechanisms and predicted influences on gaze patterns. Given the integration of rhythm and motion in these games, it is hypothesized that gaze patterns may exhibit temporal autocorrelation, where past gaze positions could predict future positions.

The data was gathered by Ainzenman et al. using an HTC Vive Pro Eye Virtual reality headset with an integrated eye tracker to capture gaze data in three-dimensional XYZ coordinates. The setup allowed for precise measurement of eye movements during gameplay. The data for each game session was stored in text files, with each file containing approximately 16,000 lines of data representing the XYZ coordinates of gaze positions, along with hit point and origin coordinates. Depth buffer values were recorded through image capture during each session to create trends for predicted fixation distances after the recording sessions.

Given the nature of the data and the research questions, the analysis primarily employed time series analytical techniques with statistical tests to explore the gaze data's underlying patterns and behaviors, deviating from traditional statistical evaluation using descriptive and inferential statistics. First, participant 1's left eye X and Y gaze data for Beat Saber and Pistol Whip were processed, formatted, and cleaned for data analysis. Only one eye's data in both games were analyzed, as patterns in the left eye would occur similarly in the right eye. This also simplified analysis. Time series plots of the X and Y coordinates of the left eye allowed for visual inspection of any apparent trends, periodicity, or stationary characteristics. To examine the presence of temporal patterns in gaze fluctuations for both games, autocorrelation function (ACF) plots were used. The autocorrelation function plot is an essential tool in time series analysis to identify whether values in a series are correlated with others at different time intervals/ lags. This is also the premise of the autoregression (AR) model, so if an ACF graph shows that a series is autocorrelated at lags, it implies the AR model may be applicable to quantify this relationship. This graph also qualitatively provides information on the data's stationarity before statistical tests and model fitting. Non-stationary time series contains a unit root and unpredictably wanders or drifts away from start points without bounds. Stationary time series are often required by modeling techniques to produce reliable and meaningful results. Gradual declines to 0 in ACF graphs indicate data stationarity. Statistical validation of stationarity utilizes the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The ADF test checks for a unit root in the data to confirm stationarity with a significant p-value. Conversely, the KPSS test assesses trends and confirms non-stationarity with a significant p-value. These tests are complementary, ensuring robust evaluation of stationarity, an important assumption of fitting autocorrelation models. Failure of one or both of these preliminary tests would demand transformation of the data through differencing consecutive observations, to account for trends in the data before fitting the time series model. This prevents biases and inaccuracies from being introduced into estimates.

After satisfaction of the assumptions of stationarity and presence of temporal patterns using ACF plots, ADF, and KPSS tests, the AR model was fitted to the data. Subsequent evaluation of residuals employed the Ljung Box Test and residual-ACF graph to evaluate whether the fit model accounted for all underlying characteristics of the data. Significant results with the Ljung Box Test applied to residuals would imply the autoregression model imperfectly fits that data, as it did not capture all patterns or correlative trends in the data. This would necessitate further refinement in the modeling approach to adequately capture all influences on the data.

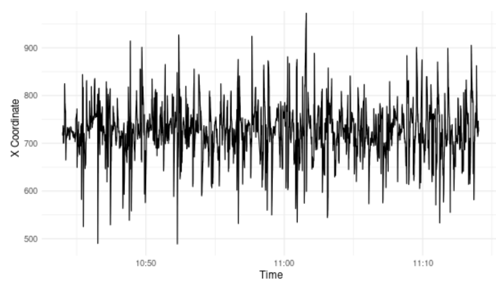
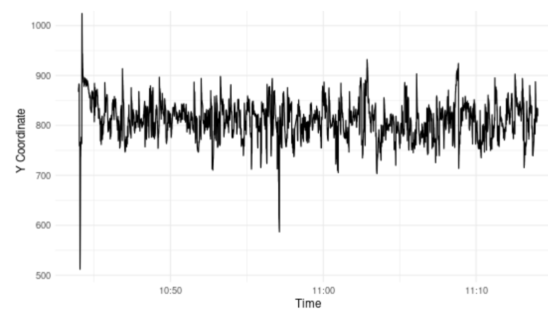
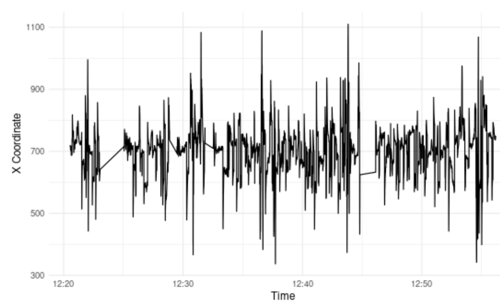
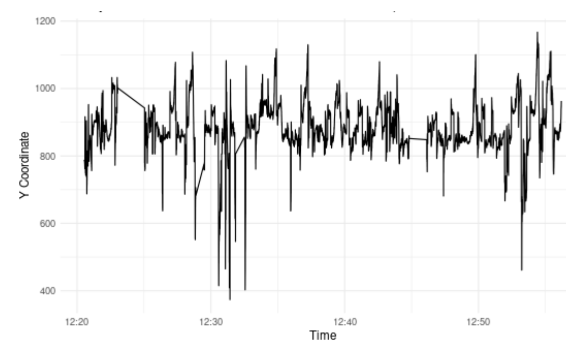
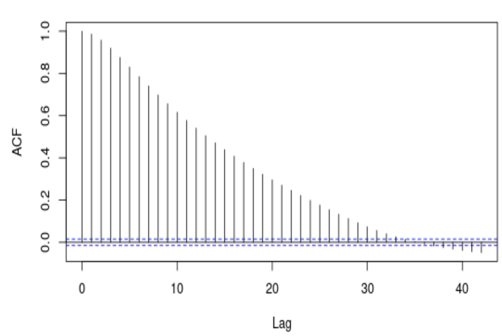
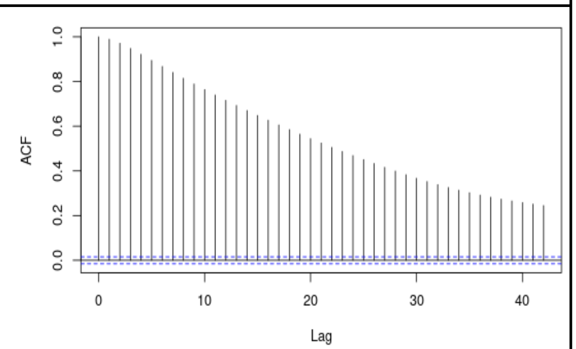
Descriptive Graphs	**X-axis Coordinates**	**Y-axis Coordinates**
Beat Saber Data		
Pistol Whip Data		

Figure 1. Line graphs showing X and Y coordinate change over time for Beat Saber and Pistol Whip

ACF Graphs	**X-axis Coordinates**	**Y-axis Coordinates**
Beat Saber Data		

Pistol Whip Data

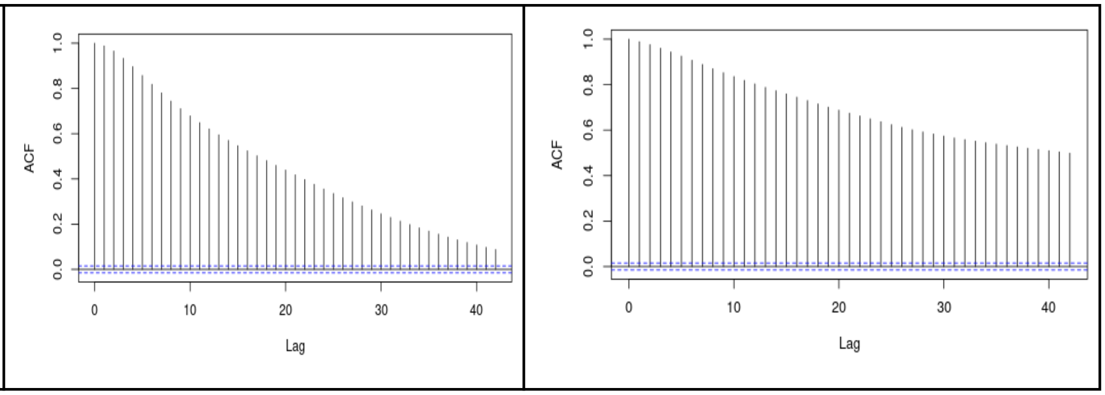


Figure 2. Autocorrelation graphs suggesting correlation in time series values and values at different time intervals.

Pistol Whip Y-Axis Residual-ACF	Beat Saber Y-Axis Residual-ACF

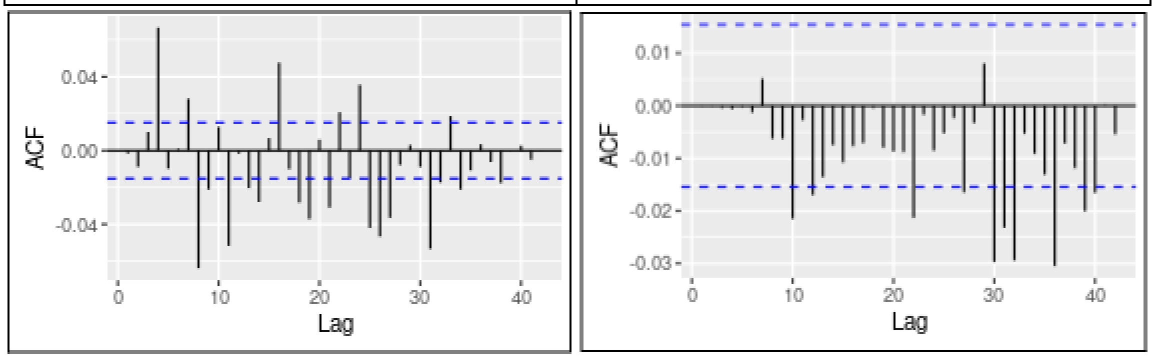


Figure 3. Autocorrelation function (ACF) plots of residuals illustrating positive autocorrelation patterns present in Pistol Whip Y-Axis residuals, and lack of positive autocorrelation patterns present in Beat Saber Y-Axis residuals.

Results section: There was temporal autocorrelation present in Beat Saber gaze fluctuation data on the Y-axis (ARIMA(5,0,0), df = 5, MAE = 2.9). The Augmented Dickey-Fuller test and Kwiatkowski-Phillips-Schmidt-Shin test reveal stationarity in the X and Y coordinate data after differencing for both games (ADF $p < 0.05$; KPSS $p > 0.05$), suggesting lack of non-stationary-characteristics. The ARIMA(5,0,0) AR model for Beat Saber's Y-coordinates fitted well, with the Ljung-Box test on residuals showing no significant autocorrelations ($p > 0.05$), indicating that the models adequately captured the underlying temporal patterns in the gaze data.

Prior to fitting an autoregressive model, assumptions regarding the data's structure and underlying characteristics had to be evaluated. Initial visualizations of coordinates against time revealed no apparent seasonality or periodic trends (**Figure 1**), prompting a deeper investigation through both qualitative and quantitative methods. Autocorrelation function plots for all data demonstrated a gradual decline towards zero (**Figure 2**), suggesting autocorrelation at various lags and justifying the choice of fitting an autoregressive model. Before fitting an AR model, the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests were applied to assess the stationarity of the time series data. For the X-coordinate data of both games, the ADF test results confirmed stationarity ($p < 0.05$), indicating no unit root was present and rejecting the null hypothesis of no stationarity. The KPSS test for the same datasets also suggested stationarity ($p > 0.05$), not allowing rejection of the null hypothesis of stationarity. This allowed continuation to fit X-axis data to the AR models. Conversely, the ADF test indicated stationarity for the Y-coordinate data ($p < 0.05$), but the KPSS test suggested non-stationarity due to detectable trends ($p < 0.05$), necessitating data differencing to stabilize variance and ensure accurate ARIMA modeling. ARIMA models were fitted to the non-differenced X-axis and differenced Y-axis datasets. The ARIMA model fitted to the differenced Y-axis data of Beat Saber achieved a good fit, evident from the low mean absolute error (MAE = 2.9), suggesting that the model's predictions were, on average, within 2.9 units of the true values. This model captured the underlying dynamics of the data. However, to prove the validity of applying the AR model to the data, analysis of the residuals takes precedence. Residual analysis using the Ljung-Box test on the Beat Saber differenced Y-axis data model residuals showed no significant autocorrelations ($p > 0.05$), indicating that the models captured the underlying patterns effectively, and that the residuals were essentially random white noise (ideal for well-fitted time series models). The Ljung-Box test applied to the 3 other model's residuals revealed significant autocorrelation in the residuals ($p < 0.05$), indicating the ARIMA model's inability to capture all characteristics influencing autocorrelation in the datasets. Autocorrelation-residual plots (**Figure 3**) were created in tandem with the Ljung-Box test. ACF values exceeding the positive confidence intervals indicate the presence of positive autocorrelation in the residuals, refuting the efficacy of the fitted ARIMA model. The Beat Saber Y-axis ACF-residuals plot (**Figure 3**) demonstrates negligible autocorrelation, further validating the ARIMA model's comprehensive capture of all underlying patterns in the data. Conversely, the residuals autocorrelation plot for the Pistol Whip Y-axis data reveals positive autocorrelation values surpassing the confidence intervals, indicating that the model may not have fully accounted for all the dynamic elements affecting the gaze behavior in this game. This discrepancy suggests the presence of additional factors or trends in the Pistol Whip data that the current ARIMA model configuration failed to incorporate, potentially requiring a more complex model or additional data preprocessing steps to achieve a similarly effective fit as observed with the Beat Saber data.

The results indicate that initial non-stationarity posed challenges; however, differencing the data effectively addressed these issues, enabling the successful application of ARIMA models that adequately captured the temporal patterns in gaze data in Beat Saber. The absence of significant autocorrelation in the residuals after modeling confirms that the Beat Saber Y-axis model was robust, and effectively accounted for the temporal dependencies within the gaze data.

This analysis substantiates the presence of temporal autocorrelation patterns in gaze fluctuations for one of the tested games, refuting the null hypothesis and supporting the alternative hypothesis that gaze movements are not random but follow discernible trends depending on the VR game. These results highlight the significant impact of game design on gaze behavior and illustrate the potential of utilizing such data to enhance user engagement and the design of VR games.

References

Dr. Karam for advising on the direction of the project and methods of analysis.

2.8 Autocorrelation | Forecasting: Principles and Practice (3rd Ed). *otexts.com*, https://otexts.com/fpp3/acf.html. Accessed 2 May 2024.

Ahmed, Ilyas. “What Are ACF and PACF Plots in Time Series Analysis?” *Medium*, 31 May 2023, https://ilyasbinsalih.medium.com/what-are-acf-and-pacf-plots-in-time-series-analysis-cb586b119c5d.

Aizenman, Avi, et al. *The Statistics of Gaze during VR Gaming*. 3, [object Object], 17 July 2022, p. 31995475230 bytes. *DOI.org (Datacite)*, https://doi.org/10.6078/D1BB16.

Aizenman, Avi M., et al. “The Statistics of Eye Movements and Binocular Disparities during VR Gaming: Implications for Headset Design.” *ACM Transactions on Graphics*, vol. 42, no. 1, Jan. 2023, p. 7:1-7:15. *ACM Digital Library*, https://doi.org/10.1145/3549529.

“Autocorrelation: What It Is, How It Works, Tests.” *Investopedia*, https://www.investopedia.com/terms/a/autocorrelation.asp. Accessed 2 May 2024.

Bobbitt, Zach. “Ljung-Box Test: Definition + Example.” *Statology*, 14 Feb. 2020, https://www.statology.org/ljung-box-test/.

Choosing the Best q and p from ACF and PACF Plots in ARMA-Type Modeling | Baeldung on Computer Science. 3 Sept. 2021, https://www.baeldung.com/cs/acf-pacf-plots-arma-modeling.

dlnB. “Answer to ‘How Do I Interpret Mean Absolute Error (MAE) or Mean Absolute Percentage Error (MAPE) in Layman Words?’” *Cross Validated*, 29 Apr. 2020, https://stats.stackexchange.com/a/463562.

G, Vijay Kumar. “Statistical Tests to Check Stationarity in Time Series.” *Analytics Vidhya*, 16 June 2021, https://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/.

Josef Perktold, et al. *Statsmodels/Statsmodels: Release 0.14.2*. v0.14.2, [object Object], 17 Apr. 2024. *DOI.org (Datacite)*, https://doi.org/10.5281/ZENODO.593847.

“MAE, MSE, and R 2 for the Daily Returns Forecast Using the ARIMA Model....” *ResearchGate*, https://www.researchgate.net/figure/MAE-MSE-and-R-2-for-the-daily-returns-forecast-using-the-ARIMA-model-The-Diebold_tbl3_361485054. Accessed 2 May 2024.

Pandian, Shanthababu. “Time Series Analysis and Forecasting | Data-Driven Insights (Updated 2024).” *Analytics Vidhya*, 23 Oct. 2021, https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/.

Tim. “Ljung Box Test: Definition.” *Statistics How To*, 7 Sept. 2018, https://www.statisticshowto.com/ljung-box-test/.

Time Series: Interpreting ACF and PACF. https://kaggle.com/code/iamleonie/time-series-interpreting-acf-and-pacf. Accessed 2 May 2024.

vasili111. “How Do I Interpret Mean Absolute Error (MAE) or Mean Absolute Percentage Error (MAPE) in Layman Words?” *Cross Validated*, 30 Apr. 2020, https://stats.stackexchange.com/q/463554.