

## USING SENSORS TO DETECT AND ANALYZE STUDENTS' ATTENTION DURING ROAD SAFETY TRAINING IN PRIMARY SCHOOL

Assist. Prof. Dr. Stefan Stavrev

Assist. Prof. Dr. Ivelina Velcheva

*Paisii Hilendarski University of Plovdiv (Bulgaria)*

**Abstract.** In this study, we use two sensors – Kinect v 2.0 and Tobii 5.0 in order to detect and measure the attention levels of primary school students during the road safety educational classes. We use custom-developed road safety serious games for conducting the experiments with students in primary schools. One of the evaluation tools is a web-based e-book – a digital representation of a traditional textbook. The other tool is a 3D virtual platform for safety on the road. Both tools are developed using Unity 3D. Finally, we compare the selected road safety educational tools, discuss their strengths and weaknesses and draw conclusions.

**Keywords:** road safety education; serious games; game-based learning; sensors

### 1. Introduction

Road safety education and training plays an important role in early child development. Road safety classes can be taught in several ways. The traditional way is by using educational videos, static presentation materials, physical objects (road signs, traffic lights) and textbooks (Papancheva & Dermendzhieva 2020); (Mihov & Dimitrov 2022). However, newer studies (Mihov, Stoitsov & Dimitrov 2022) suggest serious games can be used to train certain skills in road safety scenarios in primary schools.

In addition, those educational tools can be combined with different tracking devices (sensors) to automatically detect, track, evaluate and later analyze students' features, such as body position, movements, gaze point in 2D and 3D space. Attention in a certain road safety educational scenario can be measured by tracking player's eyes and taking note of the number of time spent on certain virtual or physical gameplay elements.

### 2. Related work

Serious games have long been used in different context, especially in education in simulation domains. They are entertaining but their main purpose is to teach the player something (Kankaanranta & Neittaanmäki 2009). Their classical origins

can be traced back to Clark C. Abt in 1970 in his book *Serious Games*. Mike Zyda (Zyda 2005) and (Sawyer & Rejeski 2002) provide a more recent formulation of the term. The implementation of serious games and natural user interface in road safety training and education has seen rise in recent years. For instance, the authors (Vidakis et al. 2015) formulate how to implement such games using natural user interface – by incorporating sensors in the game design and recognizing human gestures. However, sensors can be used not only for detecting and accepting control input, but also as an analytic tool that can benchmark correct behavioral patterns (Shen et al. 2014); (Basavaraju et al. 2019).

Using sensors to conduct real-time eye tracking has been proposed by (Khan & Lee 2019). The context of the latter research is measuring automobile's driver gaze and estimate the gaze point intersection within the physical world, in order to monitor and prevent road-related injuries. The authors use regular RGB cameras that constantly monitor the driver's eyes, send the video stream to a computer where real-time analysis occur. Kinect, on the other hand, as extensively been used as a natural user interface input device. The authors (Qingtang et al. 2015); (Szczurowski & Smith 2018); (Abdulazeez & Whittinghill 2015) and others have demonstrated how to use the Kinect sensor as a user input device for serious gaming. However, little research is done to monitor and analyze correct behavior, especially for road safety training in primary schools.

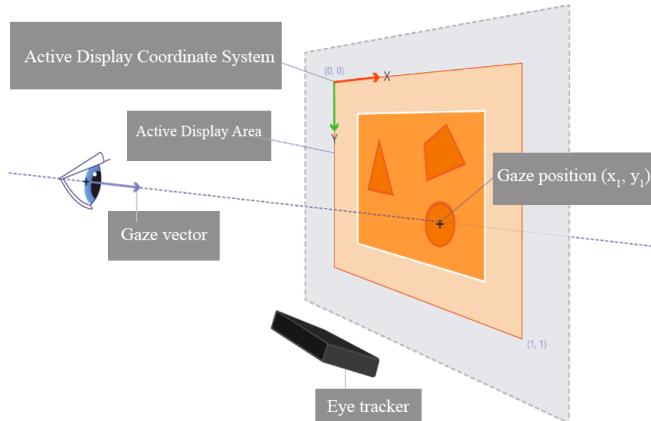
### **3. Experiment setup**

For detecting the level of attention students are paying during their road safety educational courses, we decided use two software instruments – one is a digital textbook and the other is a 3D serious game simulation environment. For our experiment, we decided to use several commercial-off-the-shelf (COTS) devices for integration and evaluation with the proposed road safety serious games. The first one is the Microsoft Kinect 2.0 sensor. It has support for serious games, targeting Windows OS. It requires a strong CPU (of at least Intel i5 quad-core with 2.2 GHz) and a USB 3.0 interface. Ideally, we would have used and the Kinect 1.0, however, that sensor is no longer officially available on the market. Kinect 2.0 is a composite device – it has two cameras – RGB and depth, and a microphone array. It can do full-body human skeletal tracking. The latter feature relies on 25 individual tracking points that combined build-up the representation of a human skeleton. By analyzing the skeletal position in certain times during gameplay, we can determine how well the player is position and oriented in the virtual and physical environments. Hence, we can determine whether he or she is positioned in the correct way by the road, whether is facing the correct direction before crossing a street, whether is paying attention to traffic lights or road signs.

Tobii sensors, on the other hand, are designed for tracking eyes and estimating gaze points. They cannot do full body skeletal tracking. Therefore, their use

is complementary to that of Kinect. We use the current 5.0 pro version of the Tobii eye tracker. The device and its application in gaze estimation, accuracy, and head-tracking capabilities has been extensively studied by (Gibaldi et al. 2017). It has the ability to track both eyes of the user in real-time. Furthermore, given the eye position, the device is capable of calculating onscreen gaze point estimation. The framework of the developer called Tobii SDK has an API that is able to read the following tracker information:

- gazeX: The x-coordinate of the estimated gazepoint on the computer screen (in pixels);
- gazeY: The y-coordinate of the estimated gazepoint on the computer screen (in pixels);
- posX: The x-coordinate of the eye pupil(s) – left / right / both pupils can be selected via property;
- posY: The y-coordinate of the eye pupil(s) – left / right / both pupils can be selected via property;
- closeTime: The time period for fixation of a particular spot on the screen;
- closeTime: The time period for closing both eyes (or eye tracking signal lost).

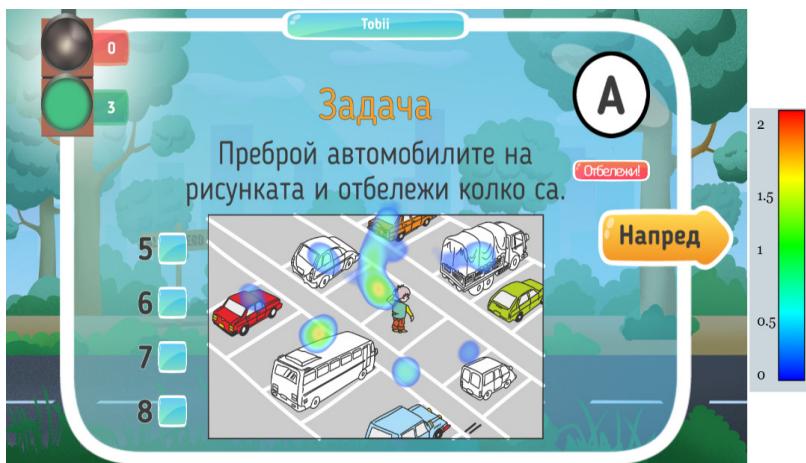


**Figure 1.** Active Display Coordinate System (ADCS), used by Tobii SDK<sup>1</sup>

Tobii SDK uses several 2D or 3D coordinate systems, and developers can choose which one to use, depending on use cases. The one used in this study outputs the screen coordinates of the gaze area (Figure 1). This is the so-called Active Display Coordinate System (ADCS).

It is a 2D coordinate system and in it, the origin (0, 0) is aligned with the top-left corner of a computer screen. Other supported coordinate systems are the User Coordinate System (3D) – with the origin centered at the eye tracker, the Track Box Coordinate System (3D) – a normalized coordinate system that creates a bounding box around the eyes and two Coordinate Systems for head-mounted displays. The

2-dimentional ADCS is preferable because it aligns with the coordinate system of the road safety serious game, the **RS E-book**. Reading and recording the *X* and *Y* gazepoint coordinates, we were able to implement a *heatmap* functionality – record the gaze point change in position on the screen and for each passing second paint a color-coded blob. The color-coding is dependent on the number of seconds the player has held his/her gaze on a certain point (Figure 2). The scale goes from 0 sec. (dark blue) to 2 sec. (dark red), as depicted in Figure 2.



**Figure 2.** Heatmap overlay of a game scenario during the RS E-book

The Tobii SDK provides the libraries necessary to access the eye tracking data in the C/C++, C#/.NET, and Unity 3D programming languages. In order to enhance the accuracy (Gibaldi et al. 2017) of the gaze point estimation, the Tobii EyeX Engine provides a native calibration procedure (TNC) to be performed before the usage of the eye tracker by a new user. We use the term “area” since the gaze estimation task is not pixel perfect but rather describes a probability gaze region. With the SDK, we are able to construct the attention heat map, as suggested in previous studies (Rigaud et al. 2016); (Manolova et al. 2021).

The first evaluation tool used with Tobii is a web-based road safety e-book, called **RS E-book**. It is a digital successor to a more traditional static textbook and is implemented as a serious game. The **RS E-book** presents to students in primary school different scenarios, related to road safety, such as where to sit in a car or motorcycle, on which side of the street to walk when there is no pavement, to cross on green light only, etc. After obtaining the gaze *X* and *Y* coordinates, an algorithm counts the number of milliseconds the player’s eyes are fixed on a certain portion of the screen. Using that information, we draw a heat map and overlay it with the web browser

view. The heat map is color-coded (Figure 2). The longer the student's gaze is fixated on a certain portion of the web-tool **RS E-book**, the redder that area becomes. In contrast, the less time is spent gazing at a screen area, the bluer it is in the heat map.

The other tool is a 3D *Virtual Environment Simulator for Educational Safety Crossing* or *V.E.S.E.S.C.* for short (Stavrev & Terzieva 2015). It was developed for road safety virtual training of children from first to fourth grade. In addition, it teaches voluntary and involuntary reactions while on the street, while crossing or waiting on the crossing walk. That tool utilizes the Kinect 2.0 composite camera to track user's hand, feet and head position and rotation, in order to control a 3-dimensional avatar, representing the player. An overlay window paints the tracked human body joints in yellow (Figure 3).



**Figure 3.** Kinect 2.0 full body skeletal tracking window (right)

While the user is looking left and right in front of a crosswalk for incoming traffic, the classroom students and teachers are able to better distinguish which way the avatar's head is rotating. This feature allows for a better understanding of the child's behavior and timely correcting any misbehavior. Brain research scientists (Knickmeyer et al. 2008); (Johnson 2005) have shown how important is learning the proper way to check for traffic in early years of one's development – 85% of the human brain develops during the first 5 years of our lives<sup>2</sup>.

#### 4. Results

We have formulated one experiment, consisting of two parts:

- Determine the correct areas at which players should be paying the most attention (i.e. baseline critical areas) for both teaching tools
- Measure the number of seconds gazing at certain area. The measurement has a quantitative part and a qualitative part.

For the web-based **RS E-book**, the attention areas are painted using a heatmap

that we presented in the previous section. In short, the longer a player focuses on certain point, the redder that are is painted. For instance, let us examine one of the training scenarios in Figure 4.



**Figure 4.** Attention levels of a student in training scenario. Focus area is mostly in front of the player but some attention is payed to the left and right vehicles

In it, the player is instructed to touch the vehicle in front of the cartoon boy. As we can see, this player's attention is mostly focused on the vehicle in front (around 1.5 – 2 seconds), but some attention (0.5 – 0.6 seconds) is directed towards the left and right vehicles. That experiment hints that this particular player seems distracted, maybe a bit hesitant about which vehicle to touch. In another scenario, a player (Figure 5) has to arrange the horizontal bars of a crosswalk.



**Figure 5.** Attention focus areas in a crosswalk scenario

As we can see, major time is spent on looking at the crosswalk and very little (0.4 – 0.5 sec.) is spent touching, dragging and positioning the horizontal zebra bars. That observation suggest correct student behavior. In the last example, we evaluate a scenario, where a player has to move an avatar on the correct (safe) path

from point *A* to point *B* (Figure 6).



**Figure 6.** Attention during a safe road evaluation between a start and a destination position

Unfortunately, in this last scenario, the player is focusing their attention on the wrong areas – such as looking at the parked car in the garage (2 seconds), looking at the groceries shop (1 second), but not paying a lot of attention on picking a safe road. (0 – 0.5 seconds). Detecting such attention discrepancies early in child road-safety training can be the first step towards taking a correcting action.

As for the other evaluation tool – the 3D virtual environment simulator *V.E.S.E.S.C.*, we measure the head position using the skeletal tracking information (provided by Kinect) and make note of the intersection point in 3D virtual space. For instance, a student is learning the basics of safely crossing a street on a cross-walk (Figure 7).



**Figure 7.** Gaze point intersection in 3D space

If the student is looking right in the middle of the crosswalk, that is correct be-

havior. However, in this instance (Figure 7) the student is looking directly ahead – the intersection with the virtual environment is near one of the trees on the other side of the street.

An important note is that the two devices that we use are proficient in different scenarios because they measure different aspects of human-computer interaction (HCI). Therefore, in Table 1 we compare Kinect 2.0 and Tobii 5.0.

**Table 1.** Comparison of technical aspects of HCI

Sensor	Kinect 2.0	Tobii 5.0
<b>Skeletal tracking</b>	yes	no
<b>Full-body tracking</b>	yes	no
<b>Eye-tracking</b>	no	yes
<b>Gaze estimation</b>	partially	yes
<b>Head tracking</b>	yes	partially
<b>Effective distance</b>	0.5 m – 8 m	0.5 m

As we can see, both devices are capable of different levels of HCI. Tobii is more suitable for close-up tracking scenarios including gaze area estimation but has no capabilities of doing full-body skeletal tracking. On the other hand, Kinect excels at full body tracking but has a very limited ability to estimate gaze points and has no dedicated eye-tracking mechanisms.

We also perform a statistical evaluation of the mean reaction time (in seconds) it takes participants to interact effectively via each sensor. Therefore, we use a one-tailed t-test. The 20 participants, involved in the experiment, first interact using the Kinect ( $M = 1.97$ ,  $SD = 0.84$ ) and then with the Tobii ( $M = 1.38$ ,  $SD = 0.46$ ). That resulted in a  $t$ -value = 2.28 and  $p$ -value = 0.013884. The result is significant at  $p < 0.05$ , i.e. Tobii is significantly better in terms of gesture detection and interaction time. However, as with any simulation, the presented tools model only the essential aspects of road safety interactions. In real-world scenarios, there are more considerations, such as attention distractions, that should be taken into account. Future efforts of current research tools can be extended into that direction.

## 5. Conclusion

In this paper, we have presented two evaluation tools, which can be used for distinguishing correct from wrong player behavior in a road-safety training. Although sensors and COTS tracking cameras are not that widespread in evaluation and analysis scenarios we have shown two instances, where those can be combined with existing serious games. Road safety education and training is crucial, especially in kindergartens and primary schools. That is why traits of correct and incorrect behavior should be detected in yearly and corrected accordingly.

### Acknowledgments

This research is supported by the Bulgarian Ministry of Education and Science under the National Program „Young Scientists and Postdoctoral Students – 2“.

### NOTES

1. Tobii SDK, Available from: <https://developer.tobiipro.com/commonconcepts/coordinatesystems.html>, [Viewed 2023-03-04].
2. National Research Council and Institute of Medicine, 2000. From Neurons to Neighborhoods: The Science of Early Childhood Development. Washington, D.C.: National Academy Press.

### BIBLIOGRAPHY

ABDULAZEEZ, A., WHITTINGHILL, D., 2015. Multiplayer Kinect Serious Games. *International Journal of Game-Based Learning*, 5(3), pp. 45 – 61.

BASAVARAJU, A., DU, J., ZHOU, F., JI, J., 2019. A Machine Learning Approach to Road Surface Anomaly Assessment Using Smartphone Sensors. *IEEE Sensors Journal*, 20(5), pp. 2635 – 2647.

GIBALDI, A., VANEGAS, M., BEX, P.J., MAIELLO, G., 2017. Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. *Behavior Research Methods*, 49, pp. 923 – 946.

JOHNSON, M.H., 2005. Sensitive periods in functional brain development: Problems and prospects. *Developmental Psychobiology*, 46(3), pp. 287 – 292.

KANKAANRANTA, M., NEITTAANMÄKI, P., 2009. *Design and Use of Serious Games*. Dordrecht: Springer.

KHAN, M., LEE, S., 2019. Gaze and Eye Tracking: Techniques and Applications in ADAS. *Sensors*, 19(24), p.5540.

KNICKMEYER, R.C., GOUTTARD, S., KANG, C., EVANS, D., WILBER, K., SMITH, J. K., GILMORE, J. H., 2008. A Structural MRI Study of Human Brain Development from Birth to 2 Years. *Journal of Neuroscience*, 28(47), pp. 12176 – 12182.

MANOLOVA, A., TONCHEV, K., NESHOV, N., CHRISTOFF, N., 2021. Human activity recognition with semantically guided graph-convolutional network. 2021 XXX International Scientific Conference Electronics (ET), 2021, pp. 1 – 4.

MIHOV, T., DIMITROV, I., 2022. The importance of STEM in primary education. Scientific works of the Union of Scientists in Bulgaria, Series V. Technics and technologies, Plovdiv 20(1), pp. 115 – 121.

MIHOV, T., STOITSOV, G., DIMITROV, I., 2022. STEM robotics in primary school. *Mathematics and Informatics*. 65(2), pp. 149 – 159.

PAPANCHEVA, R., DERMENDZHIEVA, L., 2020. Application of STEM approach in education as a factor for forming team changes and developing algorithmic thinking. *Journal „Education and Technologies“*, 11(1), pp. 57 – 61.

QINGTANG, L., YANG, W., LINJING, W., et al. 2015. Design and Implementation of a Serious Game Based on Kinect. 2015 International Conference of Educational Innovation through Technology (EITT), 2015, pp. 13 – 18.

RIGAUD, C. et al., 2016. Semi-automatic text and graphics extraction of manga using eye tracking information. *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, 2016, pp. 120 – 125.

SAWYER, B., REJESKI, D., 2002. *Serious Games: Improving Public Policy through Game-Based Learning and Simulation*. Washington, DC: Woodrow Wilson International Center for Scholars.

SHEN, Y., HERMANS, E., BAO, Q., et al. 2014. Serious Injuries: An Additional Indicator to Fatalities for Road Safety Benchmarking. *Traffic injury prevention*, 16(3), pp. 246 – 253.

STAVREV, S., TERZIEVA, T., 2015. Virtual environment simulator for educational safety crossing. *Proceedings of the 11th Annual International Conference on Computer Science and Education in Computer Science (CSECS)*, 2015, pp. 92 – 98, ISSN 1313-8624.

SZCZUROWSKI, K., SMITH, M., 2018. “Woodlands” – a Virtual Reality Serious Game Supporting Learning of Practical Road Safety Skills. *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pp. 1 – 9.

VIDAKIS, N., SYNTYCHAKIS, E., KALAFATIS, K., et al., 2015. Ludic Educational Game Creation Tool: Teaching Schoolers Road Safety. *Lecture Notes in Computer Science*, 9177.

ZYDA, M., 2005. From visual simulation to virtual reality to games. *Computer*. 38(9), pp. 25 – 32.

✉ **Dr. Stefan Stavrev, Assist. Prof.**

ORCID iD: 0000-0003-3499-101X

Web of Science Researcher iD: Q-2653-2019

Paisii Hilendarski University of Plovdiv

Faculty of Mathematics and Informatics

236, Bulgaria Blvd.

4003 Plovdiv, Bulgaria

Email: stavrev@uni-plovdiv.bg

✉ **Dr. Ivelina Velcheva, Assist. Prof.**

ORCID iD: 0000-0003-3173-4368

Paisii Hilendarski University of Plovdiv

Faculty of Mathematics and Informatics

236, Bulgaria Blvd.

4003 Plovdiv, Bulgaria

Email: ivelinavelcheva@uni-plovdiv.bg