Using Virtual Reality Environments to Predict Pedestrian Behaviour

João Filipe Costa,*† João Jacob,*† Thiago Rúbio,*† Daniel Silva,*† Henrique Lopes Cardoso,*† Sara Ferreira,*§ Rui Rodrigues,*§ Eugénio Oliveira,*† and Rosaldo J. F. Rossetti*†

*Artificial Intelligence and Computer Science Lab (LIACC)
†Department of Informatics Engineering (DEI)
‡Research Centre for Territory, Transport, and Environment (CITTA)
§Department of Civil Engineering (DEC)
Faculty of Engineering, University of Porto (FEUP)
¶INESC TEC
{up201403967, joao.jacob, reis.thiago, dcs, hlc, sara, rui.rodrigues, eco, rossetti}@fe.up.pt

Abstract—Pedestrian behaviour modelling and simulation play a fundamental role in reducing traffic risks and new policies implementation costs. However, representing human behaviour in this dynamic environment is not a trivial task and such models require an accurate representation of pedestrian behaviour. Virtual environments have been gaining notoriety as a behaviour elicitation tool, but it is still necessary to understand the validity of this technique in the context of pedestrian studies, as well as to create guidelines for its use. This work proposes a proper methodology for pedestrian behaviour elicitation using virtual reality environments in conjunction with surveys or questionnaires. The methodology focuses on gathering data about the subject, the context, and the action taken, as well as on analyzing the collected data to finally output a behavioural model. The resulting model can be used as a feedback signal to improve environment conditions for experiment iterations. A concrete implementation was built based on this methodology, serving as an example for future studies. A virtual reality traffic environment and two surveys were used as data sources for pedestrian crossing experiments. The subjects controlled a virtual avatar using an HTC Vive and were asked to traverse the distance between two points in a city. The data collected during the experiment was analyzed and used as input to a machine learning model capable of predicting pedestrian speed, taking into account their actions and perceptions. The proposed methodology allowed for the successful data gathering and its use to predict pedestrian behaviour with fairly acceptable accuracy.

Index Terms—computer simulation, microscopic pedestrian simulation, virtual reality, virtual environment, behaviour elicitation, behaviour modelling

I. INTRODUCTION

As the world population grows, it becomes essential to more effectively capitalize on urban space. Modelling and simulation are valuable tools that play a fundamental role towards this goal, by reducing risks and implementation costs [1]. However, to elicit pedestrian behaviour and, subsequently, to model pedestrians in a virtual environment are not trivial tasks. Jian et al. [2] point out that, compared to most of the other urban actors, pedestrian behaviour displays a higher degree of freedom, not heavily restrained by lanes. The high resource cost and the risk of introducing bias comprise additional hurdles to these tasks.

Pedestrian behaviour models are often classified using Lenner’s designation of microscopic, mesoscopic and macroscopic models [3]. Singular pedestrian behaviour can only be observed at the microscopic or mesoscopic scales, although in the latter behaviour is controlled by group dynamics, instead of individual decisions. Martinez-Gil et al. [4] grouped pedestrian models in five families: mechanics-based, cellular automata, stochastic, agency, and data-driven. Papadimitrou et al. [5] recognized that agent-based models are suitable to represent and simulate pedestrians at a microscopic scale, because the agents can be given certain capabilities, such as vision, cognition, and learning, which can be used by behaviours.

Behaviour elicitation is thus a challenging task of great relevance. According to Airault et al. [6], behaviour is defined as the “set of the actor’s reactions in front of what it detects of its environment”. Rossetti et al. [7] added that behaviour elicitation is not simply monitoring and generating statistics but also implementing the mechanisms needed to capture the semantics of decision processes. There are then three components that constitute behaviour:

- **Action.** The culmination of the pedestrian’s decision process. In order to move towards its goals, the pedestrian acts (or not), trying to turn the environment more favourable to itself. Studying behaviour often means to understand what leads to each action.
- **Actor.** The pedestrian itself is a key element of the behaviour. Even in the same context, different pedestrians might possess different goals and beliefs, and most likely different decision processes. Each actor processes incoming information from the environment in a different way, potentially culminating into different decisions.
- **Context.** This is not the environment itself, but its perception by the pedestrian. Decisions are not based on the real truth but on what the pedestrian believes to be the truth. The context includes not only the perception of the external environment, but also what the pedestrian perceives about itself and the present circumstances.

Traditionally, behaviour is studied through observation and
survey techniques. Observation studies completely rely on collecting data through direct observation, being susceptible to several concerns introduced by variance in observers and/or situations, erroneous sampling processes, observer bias, and inaccurate recordings [8]. Survey techniques include surveys, questionnaires, and interviews. These techniques rely on the subject, thus allowing the researcher to peek at the inner thoughts of the subjects. However, there is an inherent danger of the reported answers not being true, as people tend to portray their “ideal-self” in these scenarios [9]. Modern approaches such as video analysis through the use of computer vision algorithms are used frequently in this context [10].

Recently, virtual reality environments have been building up recognition as a tool for behaviour elicitation. Defined by Steuer [11] as a “real or simulated environment in which a perceiver experiences telepresence,” virtual reality possesses a set of characteristics which make it a suitable solution to this problem. These characteristics include: the precise measurements which attenuate errors and bias; safety, which is a major concern in the topic of transportation and mobility; relatively low costs; and the possibility of replicating scenarios and reproducing results [12]. The effectiveness of this technology has been steadily proved true, with contributions such as the SPEED framework, proposed by Almeida et al. [13], which details the behaviour elicitation process in evacuation scenarios through the use of virtual reality environments and surveys. Nevertheless, it still remains necessary to understand how this technology can be used in other scenarios, as well as to define and provide guidelines to support future studies. The metrics that define behaviour must be identified and methods to capture them must be specified.

II. Methodology

Through the use of virtual reality, users can experience environments and situations similar to those found in the real world. Creating these settings, however, is an arduous task that requires the joint work of researchers, programmers, 3D modellers, and possibly other experts. Thus, it is of relevance to attempt to maximize the results of this work when planning experiments that rely on these virtual environments. Common questions that can appear during this phase often concern topics such as the definition of the environment, the specification of the metrics to collect, the method to collect such metrics, and the process that follows the collection. While the answers to these questions often depend on the problem itself, some guidelines can be defined a priori.

A pipeline for behaviour elicitation in virtual reality, which aims to capitalize on its advantages is proposed. Figure 1 shows the different processes that constitute such a pipeline.

- **Simulation.** The starting point of the experiment is the creation of the virtual environment and its subsequent use for simulation. The environment should be populated by virtual actors, such as pedestrians and vehicles, depending on the context of the experiment. The subjects of the experiments are integrated through some kind of interface with the virtual reality simulation, where the tasks defined by the researchers, when they exist, are then performed.

- **Data Collection.** During the simulation, several mechanisms are tasked with collecting or computing the relevant metrics. While virtual reality provides the means to collect precise measures automatically, it is not suitable to extract all data, such as that pertaining to the subject. Thus, this process includes not only the mechanism present in the simulation but also additional means used to collect data.

- **Data transformation and storage.** Virtual reality is based on the use of computers. This characteristic can be capitalized to directly send the data to a database or other kinds of storage systems. If such a system supports multiple connections and remote access, multiple simulations can be stored together and performed concurrently. This process, however, encompasses the need for a data transformation step, where the computed metrics must be formatted according to the storage system, and the data collected through the various means must be fused.

- **Data analysis.** The data collected during the experiment can then be analyzed. While the nature of this step depends on the problem itself, by having stored data from multiple experiments in the same storage system, comparison studies can be performed with less effort.

- **Modelling.** While the end result of the experiment may not always be a behavioural model, in most cases it can be represented as one. If this is the case, it may be worth not only to create this model but also to implement it in the simulation, where the results can be observed in real time. By taking this extra step, the environment will be refined to a greater degree, providing better conditions for further experiments that use this environment.

Whereas this pipeline is defined as a linear process, it assumes the existence of a cycle if multiple experiments are required by an experimental design. A major concern of any controlled study is the validity of the data, as the subject’s behaviour can suffer a large influence from the conditions of the experiment. Simulation is ultimately an imitation of reality. Thus, the conscious and unconscious realization of the gap with reality constitutes a bias of unmeasured influence on the subject’s behaviour. In order to lessen this effect, the similarity between the real and virtual world should be maximized. While this process concerns several other aspects, the results of an experiment can be used to calibrate or introduce new features to the behaviour of pedestrians present in the simulation.

A. Building up the Simulation

Virtual reality is characterized by the effect of telepresence, that is, the sense of presence in an environment through the use of a medium. In order to define valid experiments with pedestrians, the environment must be as close as possible to reality. Building up this environment consists of adding several layers of rising complexity in order to achieve a realistic
The basis of the simulation is the world itself. Roads, buildings, and the sky occupy the majority of a pedestrian’s field of vision in an urban environment. These elements must be graphically believable in order to enhance the sense of immersion. An unpopulated world, however, is not complete. The next layer is composed of the actors present in the world. Any urban environment contains several entities that move around in the complex system known as the traffic flow. Such entities include vehicles, pedestrians and even animals. A pedestrian behaviour study generally aims to understand how the subjects interact with these entities and the environment. Thus, the final layer underlies the decision-making process behind the behaviour displayed by the entities. By giving them the capacity to adjust to the context, the environment can become more complex and thus respond to the subjects’ actions. Without this, interactions between the entities and with the subject will be limited to simplicity.

When developing these layers, the focus should be firstly on sight. Sight is perhaps the most important sense to simulate, and most virtual reality interfaces focus primarily on providing high-definition visual output. Thus, in general, much effort should be dedicated to delivering a visually realistic environment. The hearing and haptic senses are also key components to immersion, though there are still few devices that emulate the latter. Other senses are not as important, and very few settings can currently emulate them. As such, they should not be the focus of development unless the study requires it.

B. Behavioural Data Collection

As previously mentioned, in the scope of this work we define behaviour as a ternary relation between action, actor, and context. To fully understand a behaviour it is necessary to research all of these elements.

Regarding the action, the most basic information that can be obtained is its occurrence, for example by defining action triggers that can be used to store the specific occurrence. However, keeping track of which actions were performed might not be enough. The same action might be performed differently in different situations. For example, the action of walking can vary in speed and direction. Thus, when recording each action it might be pertinent to record this kind of data.

The actor here represents the mental state of the subject, that is, its demographic characteristics, physiological and psychological state, beliefs, goals, etc. The aim here is to shed some light upon the decision-making process then carried out by the subject. This kind of information is internal to the subject and is not directly observable through usual mechanisms that can be implemented in a virtual reality setting. In order to collect this kind of data, complementary means are usually preferred, such as surveys, questionnaires, and interviews, which are often used to get information not easily obtained through observation.

If the action is the output of the behavioural process, the context is the input: it refers to all information that the subject receives. A large part of this information is perceived from the environment, while the rest is taken from memories and thoughts of the subject. The perception process is not lossless, that is, a part of the information available in the environment is perceived incorrectly or partially. Virtual reality is suitable to collect this kind of data because all objects that can be perceived are already registered in some way. Different from other means which must detect what the subject perceives, when using virtual reality the objects are already known.

The characteristics of virtual reality make it suitable to collect data about the elements of action and context. Nevertheless, this may be insufficient regarding the actor element. Thus, it is necessary to make use of other means to complement the collected data.

The proposed methodology brings some improvements to precedent techniques. Its recursive nature provides a basis for continuous growth and revision of the simulated environment through the analysis of experimental results and recorded
feedback. Furthermore, the focus on the simulation component supports the occurrence of several parallel experiments using the same virtual environment. While earlier efforts, such as Almeida’s et al. [13], focus on designing experiments that make use of virtual reality environments, this work gives greater relevance to the creation of a virtual environment that can serve as the basis for a long series of experiments. As such, integration of this process with others is a possibility.

III. IMPLEMENTATION

In order to test this methodology, an experiment was implemented. Before starting, subjects were shown an aerial view of a virtual city, where two locations were highlighted. Figure 3 reveals the aerial view of the city, as well as the starting point (in orange) and goal (in blue). The city is fairly complete, with several different kinds of roads and intersections, including bridges, roundabouts and one-way streets. The roads are surrounded by several different buildings, and the sidewalks possess several details such as trashcans, light posts, and benches. Several virtual cars and pedestrians roam the roads of the city.

Each subject then had to traverse the distance between the two aforementioned points, commanding a first-person virtual avatar using an HTC Vive. In order to reduce external influence, the route was freely chosen by the subject. During this traversal, the data collection mechanisms constantly capture data about the subject’s behaviour, and a video of the subject field of vision in virtual reality was recorded. Subjects were also asked to answer pre-experiment and post-experiment questionnaires to complement the collected data. The former focused on demographic information about the subject, while the latter focused on understanding if any virtual reality sickness symptoms were felt and how the subjects felt about the realism of the environment. The symptoms section was based on the Virtual Reality Sickness Questionnaire [14].

A. Data Collection Mechanisms

The virtual environment is constituted by a large number of objects, most of which do not possess any value from the standpoint of collecting behaviour. These objects must be filtered out when collecting data. Each object that does have value is defined as an entity. These entities are registered on a global list and each of them keeps track of its own characteristics, updating itself when necessary. Some entities might also possess a controller which can define triggers to record the data on a server. Figure 4 shows the general structure of an entity.

An entity also provides methods to access its characteristics. These methods represent the different channels of perception which can be used to detect that entity, providing only potentially incomplete information. Thus, depending on the kind of perception, different information can be collected about the same entity. The same structure is also used for actions, as the data that characterize the action are also present in the entity. This process filters out information irrelevant to that specific perception or action.

The subject’s avatar is also considered an entity. It possesses a controller which, upon certain triggers, sends data to a local server about the actions or perceptions of the subject. Each of these actions and perceptions is associated with a verb. For this experiment, the perception verbs used were “saw,” to represent sight, and “is,” to represent spatial awareness. The action verbs considered were “walked,” “lookedAround,” and “crossed.” Different verbs are associated with different triggers. The perception verbs along with the “walked” verb represent continuous ideas, and so they are triggered at a constant rate. The remaining verbs, however, are punctual actions which possess more specific triggering conditions.

When one perception or action is triggered, the related entities are accessed, and the corresponding methods are called. If the entity supports that action’s verb, the corresponding data is returned. Then, a JSON statement is created and sent to the data storage server. The statement’s structure is based on the Experience API 1. Each statement contains a field for the actor which performs the verb, the verb itself, the objects of the verb, the time stamp, and an identifier of the simulation session where it was created. The verb field contains not only its identifier, but also an additional space denominated modifier, where the retrieved information is stored whenever it concerns the actor. Likewise, if there are objects, they can also contain the modifier field with the information relative to each one.

1https://xapi.com/overview/, last accessed: July 10, 2019
The structure is as follows:

```
{
  'actor': ...,  
  'verb': {  
    'id': ...,  
    'modifier': { ... }  
  },  
  'objects': [  
    {  
      'id': ...,  
      'modifier': { ... }  
    }  
  ],  
  'timestamp': ...,  
  'sessionId': ...
}
```

Each kind of entity has its own relevant metrics, which are assessed by different verbs. Table I contains the metrics of each entity and the verbs which are used to access them.

### IV. RESULTS

The experiment was performed with fifteen subjects aged between 16 and 30 years (mean age = 23.2 years, standard deviation = 3.47). Six subjects were female, while the remaining nine were male. All but one of them possessed a driver's license. Each subject participated in two rounds of experiments, with an average duration of eight minutes and thirteen seconds (standard deviation = 2m05s). In total, four hours, six minutes, and fifteen seconds of video footage were recorded.

From the results of the post-experiment questionnaire, it was found that experiment aroused some minor symptoms, (final mean score of 14.8 out of 100). Table II contains the mean and standard deviation values for each symptom. As it can be observed, the main reported symptom was fatigue. This can be attributed to the avatar's control method, which used arms movement to calculate the speed in the virtual world. While the action was easy to perform, maintaining it for more than eight minutes per experiment could be considered strenuous.

In terms of realism, the subjects found the environment fairly acceptable (mean = 3.9 out of 5; standard deviation = 0.52). When asked about the realism of certain components, pedestrians revealed greater acceptance for the streets, while vehicles were reported as relatively less realistic. The reported causes included an improper 3D model and unrealistic behaviours. Table III contains the scores for each component.

Using the collected data, an exploratory analysis was performed aiming to obtain a prediction model for pedestrian speed during a crossing event. A crossing event was defined as the segment of time where a subject moved from a sidewalk to another, passing through the road. Additionally, the preceding five seconds before the crossing started were also taken into account, as the impact of the event during this time was also considered significant. Using this definition, all the crossing events were extracted. For each event, data concerning speed, direction which the subject is facing, as well as if it sees cars, were extracted and transformed into time-series containing one record each half of a second.

These time series were uploaded to Rapidminer, a data science-oriented software. Using this software, a model that utilized the current and previous values of the aforementioned time-series to predict the current speed was built. The resulting model presented a root means squared error of 0.207. As seen in Figure 5, which contains an application of the model to

![Image](https://rapidminer.com/), last accessed: July 10, 2019

---

| TABLE I | METRICS COLLECTED BY ENTITY AND VERB. |
|---------|--------------------------------------|
| Entity  | Verb | Metric                        |
| Car     | saw  | Position                      |
|         |      | Rotation                       |
| Crosswalk | saw | Position                      |
|         |      | Rotation                       |
| Pedestrian | saw/is | Current activity             |
|         |      | Apparent age                  |
|         |      | Apparent Gender               |
| Pedestrian (Cockpit only) | lookedAround | Direction                 |
| Traffic Lights | saw | Light color for pedestrians   |
| Weather | saw  | State                         |
|         |      | Time                          |

| TABLE II | VRSQ SYMPTOMS SCORES (IN A FOUR-POINT LIKERT SCALE). |
|-----------|-----------------------------------------------------|
| VRSQ Symptom | Lower Quartile | Mean | Top Quartile |
| General discomfort | 0.0 | 0.5  | 1.0 |
| Fatigue | 0.5 | 1.1  | 2.0 |
| Eyestrain | 0.0 | 0.5  | 1.0 |
| Difficulty focusing | 0.0 | 0.4  | 1.0 |
| Headache | 0.0 | 0.2  | 0.0 |
| Fullness of head | 0.0 | 0.4  | 1.0 |
| Blurred vision | 0.0 | 0.5  | 1.0 |
| Dizzy (eyes closed) | 0.0 | 0.0  | 0.0 |
| Vertigo | 0.0 | 0.2  | 0.0 |

| TABLE III | REALISM SCORES (IN A FIVE-POINT LIKERT SCALE). |
|-----------|-------------------------------------------------|
| Element   | Lower Quartile | Mean | Top Quartile |
| General discomfort | 3.0 | 3.7  | 4.5 |
| Fatigue | 4.0 | 4.5  | 5.0 |
| Eyestrain | 2.0 | 2.8  | 4.0 |
| Difficulty focusing | 3.0 | 3.7  | 4.0 |
| General | 4.0 | 3.9  | 4.0 |

---
potential advantages to understanding pedestrian behaviour. Overall, the next steps rely on further experimentation, both to consolidate this work and make the methodology more robust.

ACKNOWLEDGMENT

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 723386.

Authors are grateful to the SIMUSAFE Consortium’s members for their invaluable comments and fruitful discussions throughout the course of the SIMUSAFE Project (http://simusafe.eu/), within which this work has been developed.

REFERENCES

[1] M. Batty, “Urban Modeling,” International Encyclopedia of Human Geography, Oxford, UK: Elsevier, 2009.
[2] L. Jian, Y. Lizhong, and Z. Daolang, “Simulation of bi-direction pedestrian movement in corridor,” Physica A: Statistical Mechanics and its Applications, vol. 354, no. 1-4, pp. 619–628, aug 2005.
[3] A. Lerner, Y. Chrysanthis, and D. Lischinski, “Crowds by example,” Computer Graphics Forum, vol. 26, no. 3, pp. 655–664, sep 2007.
[4] F. Martinez-Gil, M. Lozano, I. García-Fernández, and F. Fernández, “Modeling, Evaluation, and Scale on Artificial Pedestrians,” ACM Computing Surveys, vol. 50, no. 5, pp. 1–35, 2017.
[5] E. Papadimitriou, G. Yannis, and J. Golas, “A critical assessment of pedestrian behaviour models,” Transportation Research Part F: Traffic Psychology and Behaviour, vol. 12, no. 3, pp. 242–255, may 2009.
[6] V. Ayrault, S. Espié, C. Lattaud, and JM, “Interaction between pedestrians and their environment when road-crossing: A behavioural approach,” in Proceedings of the 24th Urban Data Management Symposium, E. Fendel and M. Rumor, Eds., Chioggia, 27-29 Oct 2004.
[7] R. J. F. Rossetti, J. E. Almeida, Z. Kokkinogenis, and J. Goncalves, “Playing transportation seriously: Applications of serious games to artificial transportation systems,” IEEE Intelligent Systems, vol. 28, no. 4, pp. 107–113, 2013.
[8] H. M. Hughes and S. N. Haynes, “Structured laboratory observation in the behavioral assessment of parent-child interactions: A methodological critique,” Behavior Therapy, vol. 9, no. 3, pp. 428–447, jun 1978.
[9] A. Millonig, N. Brändle, M. Ray, D. Bauer, and S. C. Van der Spek, “Pedestrian Behaviour Monitoring: Methods and Experiences,” in Behaviour Monitoring and Interpretation - BMI, B. Gottfried and H. Aghajan, Eds. Amsterdam: IOS Press, 2009, vol. 3, pp. 11 – 42.
[10] P. Afsar, P. Cortez, and H. Santos, “Automatic visual detection of human behavior: A review from 2000 to 2014,” Expert Systems with Applications, vol. 42, no. 20, pp. 6935–6956, Nov 2015.
[11] J. Steuer, “Defining Virtual Reality: Dimensions Determining Telepresence,” Journal of Communication, vol. 42, no. 4, pp. 73–93, Dec 1992.
[12] M. Kinateder, E. Ronchi, D. Nilsson, M. Kobs, M. Müller, P. Pauli, and A. Mühlsberger, “Virtual Reality for Fire Evacuation Research,” in 2014 Federated Conference on Computer Science and Information Systems, Warsaw, Poland, 7-10 Sept 2014, pp. 313–321.
[13] J. E. Almeida, J. T. P. N. Jacob, B. M. Faria, R. J. F. Rossetti, and A. L. Coelho, “Serious Games for the Elicitation of Way-finding Behaviours in Emergency Situations,” in 2014 9th Iberian Conference on Information Systems and Technologies (CISTI), Barcelona, Spain, 18-21 Jun 2014, pp. 1–7.
[14] H. K. Kim, J. Park, Y. Choi, and M. Choe, “Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment,” Applied Ergonomics, vol. 69, pp. 66–73, May 2018.