QUANTIFYING USER PHYSICAL ACTIVITY IN A VIRTUAL ENVIRONMENT: THE CASE OF SyncVR Medical

SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE

Bonaventura Francesco Pacileo
12634417

MASTER INFORMATION STUDIES
DATA SCIENCE
FACULTY OF SCIENCE
UNIVERSITY OF AMSTERDAM

2020-08-26

<table>
<thead>
<tr>
<th>Title, Name</th>
<th>Internal Supervisor</th>
<th>External Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robert Belleman</td>
<td>Freek Van Polen</td>
<td></td>
</tr>
<tr>
<td>UvA</td>
<td>SyncVR Medical</td>
<td></td>
</tr>
<tr>
<td><a href="mailto:r.g.belleman@uva.nl">r.g.belleman@uva.nl</a></td>
<td><a href="mailto:freek@syncvr.tech">freek@syncvr.tech</a></td>
<td></td>
</tr>
</tbody>
</table>

Universiteit van Amsterdam
Quantifying user physical activity in a virtual environment: the case of SyncVR Medical

Bonaventura Pacileo
University of Amsterdam

ABSTRACT

The clinical effectiveness of Virtual Reality-based (VR) treatments could be enhanced by insights derived from VR devices sensors. The proposed system effectively complements VR-based treatments by tracking and displaying patient progress on a sleek dashboard. It is developed using a modern, scalable and modular cloud architecture. The human activity recognition component, trained on VR devices’ sensor data, is accurate at predicting user gestures in VR and is constantly re-trained as new labeled data becomes available.

1 INTRODUCTION

Virtual reality (VR) is steadily getting traction in healthcare, re-inventing services both at the clinical point of care (inpatient) and in telemedicine solutions (outpatient) [16–18]. In parallel, the application of machine learning models to body-worn sensor data for human (physical) activity recognition (HAR) has garnered a lot of interest in fitness and healthcare solutions [3, 7, 10, 13, 14], albeit little attention has been devoted to the potential of sensor data from VR devices. Finally, little attention has been devoted to how end-to-end data pipelines could improve the efficiency of VR-based therapies by accurately measuring and visualising patients’ physical activity and adherence to treatment.

In the last decade, research has evaluated - under experimental conditions - the effectiveness of VR-based treatments for rehabilitation therapy [9, 16–18, 21]. Moreover, research has been conducted on developing physiotherapy metrics based on of patients’ activity data and their visualisation [2, 15]. In fact, these visualisations could i) engage and motivate patients [15], ii) foster research into developing comparable and homogeneous metrics and iii) enhance awareness about the clinical usefulness of sensor data from VR devices [20]. These strands of research have paved the way to investigate how VR-based rehabilitation solutions could harness visualisation tools and insight derived from sensor data to improve their clinical effectiveness and promote their uptake.

Developing VR-based treatments that incorporate sensor-based clinical metrics, embedded (machine learning) analytics and visualisation tools amounts to building a complex data science system. To go about it, it is necessary to build and connect the following functional components:

- VR applications and telemetry;
- Patient-centered metrics;
- Data extraction, processing and storage;
- Data labeling, model training and deployment;
- Data visualisation tools;
- Continuous integration (CI) tools.

Existing solutions did not undertake a comprehensive approach and have instead focused on a subset of the required components [1–3]. This is possibly linked to the relative immaturity of health tech solutions, as they sit at the intersection of rapidly evolving fields. It is thus argued that deploying a VR application embedding sophisticated telemetry with the aim of improving patients’ rehabilitation journey provides a solid use case to challenge the status quo and to develop a solution that could combine all those components.

The proposed solution is thus a data science system that continuously records, processes and visualises data produced by user activity in a VR rehabilitation application developed by the company SyncVR Medical. This VR application has been developed to support patients rehabilitating from chronic lower back pain (CLBP). As shown in Figure 1, this system leverages sensor data from VR devices to develop clinical indicators that become part of a reporting dashboard. It is argued that this solution enhances the user experience of such VR-based treatments and improves their effectiveness.

The design and development of this solution provides compelling evidence that

1 Activity-based indicators such as total time, number of actions and maximum range of motion allow physiotherapists to track patient rehabilitation from CLBP;
2 Patients benefit from a complementary motion capture (mo-cap) tool that plays back user actions;
3 A deep learning model with Long Short-Term Memory layers accurately predicts clinically relevant user actions (topping 70% accuracy);
4 A dashboard powered by these indicators enhances the effectiveness of VR-based treatments as reported by physiotherapists and patients;
5 To fulfill its use case, this dashboard needs to be updated after every user session.

1https://syncvr.tech
Section 6 presents a comprehensive list of the terms used in the article.

Figure 1: Conceptual model

2 RELATED WORK

The use of VR for medical applications, especially in physical rehabilitation, has long been investigated as it affords precise stimulus control, immersive, dynamic, 3D environments and performance tracking and recording [2, 16]. Moreover, VR could bring several advantages to the patients undergoing treatment. In fact, VR enhances the rehabilitation of patients [5, 22], while ensuring error-free learning [16], reducing levels of distress [18] and minimising kinesophobia [19].

Objective metrics have to be defined in order to measure task performance, progression in rehabilitation [16, 23] and improvements in patients activities of daily life (ADL) [2]. However, the development of such metrics based on the user behaviour during VR sessions is non-trivial, as biomechanical, physiological and technical constraints have to be to taken into account [17, 21].

Human activity recognition (HAR) could be conducted by mining sensor data from mobile devices (mostly accelerometers and gyroscopes) [10, 11]. From a machine learning standpoint, this ties strongly to the classification of physical activities during VR physiotherapy sessions. However, limitations have been identified in the deployment at scale of such HAR models [12], possibly due to heterogeneity in the experimental setups in terms of data collection methods, sample structure, hardware specs and data processing and analysis techniques [11]. In this regard, deep learning models such as Long Short-Term Memory (LSTM) could be successfully applied to HAR classification problems and possibly overcome some of the above limitations [14].

HAR relates strongly to the emerging field of digital phenotyping. Digital phenotyping has been defined as "[the] moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices" [13]. In fact, HAR leverages sensor data from mobile devices to measure human activities. From a medical standpoint, this represents a cultural shift in that measurements can be carried out in naturalistic settings and without the incremental expenses and personal burden associated with additional specialized equipment [4, 7, 13, 20]. This is relevant, as accelerometer data could be used broadly to quantify physical mobility patterns and metabolic expenditure in a range of patients while tracking the full spectrum of their disease from diagnosis, to treatment and to chronic disease management unobtrusively [7].

Finally, the visualisations of such metrics must be developed with a user-centered approach [2, 15]. Such visualisations could include include both performance (i.e. movement topology) and result (i.e. movement outcome) feedback [5]. In fact, physiotherapists could use a dashboard that shows objective patient data before or during consultations to assess patient engagement and their performance over time [15]. Moreover, an intuitive visualization tool affords a more active involvement of the patients in the rehabilitation process and an increased probability of effective transfer training [15]. In fact, physiotherapists have to identify the extent to which a specific skill or a general familiarity with the training context is being transferred [23].

3 IMPLEMENTATION

In production, the solution operates according to Figure 2. In particular:

- Whenever a user completes a session on the VR application, the application will send the data to the Data Storage;
- That action triggers the Data Preparation, including calling the deep learning model to predict user actions;
- The resulting data is then available to patients and physiotherapists on the Reporting Dashboard.

Before delving into the technical specifications of the solution, the underlying technology is briefly introduced.

Technology

The Oculus Quest[^2] is an integrated computer unit and head-mounted display (HMD) with separate hand controllers. HMD and hand controllers are called markers, and they afford six channels - rotational and translational tracking in 3D space - of movement data.

[^2]: oculus.com
The VR rehabilitation application was originally developed in Unity\(^3\), and runs at 72 frames per second (fps).

Cloud solutions, such as Google Cloud Platform\(^4\) (GCP), offer a wide range of data engineering and data science tools on a metered pay-as-you-go basis. The following services have been used to develop this solution:

- Storage and database (Cloud Storage buckets and Cloud Firestore);
- Virtual machines (Compute Engine);
- Machine-learning tools (AI Hub and TensorFlow);
- Serverless event-driven functions (Cloud Functions);
- Continuous Integration tools (Cloud Build and Cloud Run);
- Reporting tools (Data Studio).

Finally, all code has been written in Python 3.7.

**Clinical indicators**

The development of the clinical indicators was informed by a stakeholder research with several physiotherapists and patients\(^2, 15\). The Oculus Quest HMD affords an extremely accurate tracking of the user head, which was used to establish the following set of metrics:

- **Activity recognition**: left/right rotations, bends, extensions, flexions and standing;
- **Head movement analysis**: maximum range of motion of the neck across activities\[^22\];
- **Fuel score**: a composite indicator of time and effort as defined by the metabolic equivalent of task (MET)\(^6\);
- **Head position analysis**: a heatmap of the user head positions based on the kernel density of the sensor positional data.

Moreover, a mo-cap tool was added to allow patients and physiotherapists to play back user actions. The tool displays in a 3D animated environment a humanoid figure that repeats the user actions performed during the VR-based treatment session. Besides giving physiotherapists an holistic view of patient progress over time, it could also allow patients to maximise transfer training from the therapy\(^9, 19\).

**Solution**

From a software architecture standpoint, the solution consists of a relatively scalable and loosely coupled architecture; once the components are all connected into a pipeline, the system can timely derive analytics of the data recorded on the VR device(s). Figure 2 offers a glimpse into the pipeline architecture.

The **Data Storage** is pivotal to this architecture: it combines a Cloud Storage Bucket for heavy data (e.g. the movement data logs) and a Firebase NoSQL database for the activities directly recorded by the VR application (e.g. the interactions between users and the virtual environment). The VR physiotherapy application automatically stores movement data to the bucket at the end of each session.

To predict the clinically relevant user actions defined above, a LSTM model is used. This deep learning model is especially suited for sequential data\(^14\) and the raw data fed to it does not require any preprocessing that could potentially introduce bias. While at runtime the system makes inferences on the new data, the underlying model could be...
continuously trained as new labelled data is collected.

As shown in Section 4, training data requires manual labelling. This process of annotating data is time-intensive but greatly contributes to improving the accuracy of the model. Figure 3 shows how the training process could be automated so that every time new labelled data is collected, the deep learning model is automatically re-trained and, provided that the accuracy increases against a given test set, the new model version is deployed to production.

A Cloud Storage bucket stores the accuracy metrics of the live model and the test dataset against which the model trained using also the newly labelled data is compared.

The Data Preparation is the automated service dedicated to listening to writing events of the VR physiotherapy application into the Data Storage. It relies on a Cloud Function to process multiple data files in parallel. In fact, Cloud Functions are a function-as-a-service (FaaS) tool which provides a platform to run applications without requiring to build or maintain infra. Thus, a larger number of requests will be met by the platform by automatically provisioning more infra, with no manual intervention required. The Cloud Function calls the serving LSTM model to make inferences on user movements. Concurrently, the other physical activity metrics are computed by the same function. Final results are stored, ready to be consumed by the reporting tool.

The Reporting Dashboard is the third main service of the solution. It is hosted on Google Data Studio and it listens to the bucket containing the processed results via an API. This tool allows patients and physiotherapists alike to review and analyse user actions and to monitor their progress.

In production, this system requires to be constantly updated as new pipeline components are added (e.g. new indicators are introduced) or existing ones are improved (e.g. Python code is optimised). From a DevOps perspective, it is thus required to have a pipeline in place to build, test and deploy newer versions of these services, i.e. a continuous integration (CI) tool. Algorithm 1 conceptually presents the different steps of how such pipeline is currently operating.

### Software development

This solution shall be considered part of the product development of SyncVR Medical. As such, the development followed the same iterative approach of other SyncVR Medical products, inspired by the Agile\(^2\) methodology. In fact, sprints of two weeks were followed by demo sessions where feedback was garnered and stakeholder were contacted for rapid iteration and testing. All code is available at this Git repository.

## 4 EVALUATION

### Feasibility

The Data Storage, the Data Preparation and the Reporting Dashboard services, event-triggered functions and synchronous connectors are built atop existing cloud-computing services. GCP was chosen for its stronger focus on deep learning and to leverage pre-existing SyncVR Medical cloud infra on GCP for mobile device management.

Event-triggered functions such as Google Cloud Functions naturally afford scalability since they are serverless and do not require any manual provision of infra. Thus the system can handle spikes in demand and only suffers from a little latency due to cold invocations. However, this latency is negligible and the dashboard on average shows new results in less than two minutes. This is consistent with the functional requirements of the dashboard, since physiotherapists reported that at the end of every session patients need to take off all the VR equipment and move back to the office room before reviewing the latest results. These operations happen in a similar timeframe as the dashboard computing time. Moreover, Cloud Functions efficiently manage asynchronous jobs - like the VR-based rehabilitation sessions.

Services are organised around a microservice architecture, so that breaks will have a reduced impact on user experience. This is a key requirement of live systems where on the one end users continuously create data through their interactions with the VR devices and on the other end they consume

---

\(^2\)https://www.agilealliance.org
data-driven results presented by the front-end tools. As side benefits, this modular architecture allows to add faster computing instances should circumstances require less latency and to add additional services to the existing infra. For instance, during the development of the system, the Reporting Dashboard was complemented by a daily reporting email tool, as external circumstances required to scale up the outpatient capabilities of the system. From a cloud engineering standpoint, this addition did not require any overhead besides connecting the email tool to the results bucket.

From an economic standpoint, the use of a public cloud and the adoption of event-triggered functions are a great advantage. On the development side, they allow rapid experimentation and iteration with reduced costs especially in combination with CI tools such as Algorithm 1. On the production side, they match costs to effective use on a metered pay-as-you-go basis plus naturally accommodating spikes in demand.

**Experimental results**

Human activity recognition (HAR) using sensor data from mobile devices has experienced widespread adoption in recent years thanks to the application of machine learning models to quantify user physical activity in naturalistic settings [1, 3, 7, 10, 11, 13, 14]. In the following sections, the overall design choices of the HAR pipeline are evaluated first and then the different components are, including the deep learning model.

*Human activity recognition topology.* From a theoretical standpoint, the system could take several configurations in order to meet application and user requirements in terms of environment, target, recognition, periodicity and statefulness [3]. The current solution could be evaluated along these five dimensions as shown in Table 1. In fact these five dimensions present the five main challenges of a HAR system from a design and development perspective. The rationale presented for every dimension explains how the use case at hand required to choose a specific feature over its alternatives. For instance, the system does not require to update results in real-time, thus an offline execution was favoured.

*Data Collection.* The deployment of the VR physiotherapy application preceded the development of this analytics suite. As such, the data collection was pull-driven, in that data was collected whenever users were using the VR physiotherapy application in inpatient or outpatient scenarios. At a later stage, the movement data stored on the VR devices was exported to the bucket and since then all subsequent data was automatically streamed at the end of every session from the VR device to the bucket. Additionally, further data was collected and users were video-recorded for label annotation purposes.

For privacy issues, no sensor data could be linked back to patients directly, so it is not possible to ascertain exactly how many distinct users are part of the sample. However, a lower bound could be provided by the annotated samples, which came from 5 different users.

*Ground truth annotation.* The manual annotation is a well-recognised challenge in HAR tasks and could be implemented
in different ways [3]. As shown in Figure 3, in the current system it was achieved by i) decoupling data generation from event labeling (i.e. filming users while using the VR device) and subsequently ii) presenting video and data during labeling with timestamps (to minimise errors) [8].

Class imbalance. However, this annotation approach exacerbates class imbalance, a recurring issue in HAR tasks [3, 10, 14]. In fact, the NULL class, i.e. standing, accounts for more than 50% sample and had to be removed before analysis.

Performance evaluation. The model specification and the related definition of the sequence length of the raw movement data were optimised with a manual hyperparameter search. This drew extensively from related research [14], albeit this is the first study to investigate human activity recognition at the gesture level from VR sensor data.

The model was trained using TensorFlow on a Google Compute Engine instance with 4 vCPUs and 15 GB RAM. Training the model takes around 430 billable seconds.

The LSTM model, that could be improved in the future as more data is collected, reaches a validation accuracy of 70% after 50 epochs. As a reference, the model, with relatively limited fine-tuning and training time, performs 7x better than a random guess (10% with 10 classes). The 10 classes are presented in Table 2.

<table>
<thead>
<tr>
<th>Type</th>
<th>Feature</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Execution</td>
<td>Offline</td>
<td>Results are reviewed after the session</td>
</tr>
<tr>
<td>Generalisation</td>
<td>User indepen-</td>
<td>The focus is the action and not the user</td>
</tr>
<tr>
<td></td>
<td>dent</td>
<td></td>
</tr>
<tr>
<td>Recognition</td>
<td>Continuous</td>
<td>Actions take place in the streaming data</td>
</tr>
<tr>
<td>Activities</td>
<td>Sporadic</td>
<td>Actions do not show periodicity but are interspersed</td>
</tr>
<tr>
<td>System model</td>
<td>Stateless</td>
<td>Specific sensor signals are used</td>
</tr>
</tbody>
</table>

Table 1: Human Activity Recognition

Granularity. Another related challenge is the definition of the physical activities of interest. Current research [1, 10] has focused on the action level (e.g. walking vs running) whereas the approach taken here is to recognise specific gestures (e.g. head bending vs flexing). This issue was addressed in the ground truth annotation. In fact, besides annotating data from regular user sessions, multiple datasets were created by one user performing a succession of the specific actions, that well later carefully annotated. This data was later fed to the deep learning to train it at recognising such low-level movements.

Performance evaluation. The model specification and the related definition of the sequence length of the raw movement data were optimised with a manual hyperparameter search. This drew extensively from related research [14], albeit this is the first study to investigate human activity recognition at the gesture level from VR sensor data.

The model was trained using TensorFlow on a Google Compute Engine instance with 4 vCPUs and 15 GB RAM. Training the model takes around 430 billable seconds.

The LSTM model, that could be improved in the future as more data is collected, reaches a validation accuracy of 70% after 50 epochs. As a reference, the model, with relatively limited fine-tuning and training time, performs 7x better than a random guess (10% with 10 classes). The 10 classes are presented in Table 2.

<table>
<thead>
<tr>
<th>Type</th>
<th>Feature</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left rotation</td>
<td>Right rotation</td>
<td></td>
</tr>
<tr>
<td>Left bend</td>
<td>Right bend</td>
<td></td>
</tr>
<tr>
<td>3D Left flexion</td>
<td>3D Right flexion</td>
<td></td>
</tr>
<tr>
<td>3D Left extension</td>
<td>3D Right extension</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>Flexion</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Model classes

The model configuration has the following features:

- Sensor data (six channels) was sampled into 144 frame-long sequences (i.e. 2 seconds with 72 fps);
• A LSTM layer with 32 hidden units, l2 regularizer and tanh activation;
• A second LSTM layer with 32 hidden units, l2 regularizer and tanh activation;
• A Dense layer with 10 classes and softmax activation.

User testing. The main motivation for developing this solution was to provide physiotherapists and patients with a simple yet effective way of supporting patient rehabilitation by visualising user performance over time and tracking their progress.

As a consequence, multiple physiotherapists and patients have been interviewed to understand the type of information they needed and the way it could be best conveyed. Figure 4 offers a glimpse into the Reporting Dashboard, whose development has indeed been heavily influenced by this stakeholder research.

All interviews followed the same structure to ensure consistency in the derived insights. The format was semi-structured, and a designated area of focus was defined at the beginning of the interview. In a first phase, the focus was on the type and format of information physiotherapists and patients would need to see in order to understand patient performance over time. In a second phase, as functional requirements were then clearly defined, a second round of interviews focused on different mockups of the dashboard tool to further refine the product and enhance user experience. Overall, three patients and nine physiotherapists were interviewed.

The Reporting Dashboard is composed of the following elements, that consistently match existing research on visualisation tools for rehabilitation [2, 15]:

- **Activity recognition**: the actions reported are the ones predicted by the LSTM model. It helps breakdown the activities and detect anomalies;
- **Head movement analysis**: supports tracking the progress over time, as the maximum range of motion is displayed for the different actions;
- **Fuel score**: gamifies the experience for patients while ensuring they regularly exercise;
- **Head position analysis**: the heatmap helps flagging instances of restricted movement areas, for instance due to kinesophobia.
- **After action review**: while new in the literature, the mo-cap is the component that all stakeholders identify as the killer feature.

More specifically, the use case for this system was to support patients rehabilitating from CLBP. The VR-based treatment focuses on exercising specific head movements, getting acquainted to (re-)doing those movements effortlessly. These movements often constitute the basis of activities of daily life (ADL) and are thus the target of these treatments.

The physiotherapists will find in the Reporting Dashboard objective measures of both movement topology (**Activity recognition**) and movement outcome (**Head movement analysis**) plus fine-grained information on possible areas of restricted movement (**Head position analysis**). In outpatient scenarios, the **After action review** allows them to observe patient behaviour and performance over time.

The Reporting Dashboard helps patients by on the one hand engaging and motivating them and on the other one by supporting transfer training while reducing distress and kinesophobia. To fulfill the first goal, patients can view on the front page a recap of their recent activity, and on the second one a **Fuel score** tool that enhances user engagement [20]. As per transfer training and minimising kinesophobia, the **After action review** tool allows patients to review their actual movements and reflect on their physical capabilities. While this seems trivial, it has been reported as a very useful tool during the interviews and it is backed by existing research [16, 19].

5 DISCUSSION

The integration of sensor-based clinical metrics into VR-based treatments must be addressed with a comprehensive approach, as the different required components all present their unique challenges. However, it is shown that i) this integration offers considerable advantages to VR-based therapies and ii) that it is feasible to develop such solution. A cloud infra with several microservices offers a viable way forward, provided that the clinical indicators and their visualisation tools are developed as part of a wider stakeholder research.

The proposed solution is feasible, well-performing and most importantly it addresses a concrete use case. Future work should pursue two main avenues: i) assembling quality publicly available VR sensors datasets, to provide benchmarks to the research community and ii) expanding the current set of available clinical indicators both for CLBP and other related diseases. From a technological standpoint, future iterations could also include more advanced analytics pipelines, to deal with exponential increases in data volume, variety or velocity. In fact, different use cases might require real-time execution of HAR, which would not be possible with the current architecture.
6 GLOSSARY

The research presented in this article cuts across multiple disciplines and thus employs a wide range of terms and acronyms both in the body text and in the figures.

A short definition of these concepts is presented below:

• Continuous Integration (CI): process of automating the build and testing of code every time changes are committed to the version control system (VCS).
• DevOps: A software engineering culture that aims at making software development, testing and deployment process as frequent, reliable and fast as possible.
• Kernel density: a data smoothing technique to estimate the probability density function of a random variable.
• Long Short-Term Memory (LSTM): a recurrent neural network (RNN) architecture composed of a memory cell and three gates that regulate the information flow, an input gate, an output gate and a forget gate.
• Metabolic equivalent of task (MET): the objective measure of the rate at which a person expends energy while performing some specific physical activity compared to sitting quietly relative to their mass.
• Motion capture (mo-cap): recording actions of human actors in physical spaces, and using that information to animate digital characters in 3D computer animation.
• Production (PROD): the production environment is the live environment that users directly interact with. In the software development lifecycle, code is created in the development (DEV) environment, then sent to testing and finally deployed to PROD.
• Unit testing: a set of automated tests to ensure that different components of an application (known as the "unit") meet its design and behaves as intended.

7 ACKNOWLEDGEMENTS

I am grateful to Freek Van Polen and to the entire SyncVR Medical team for their constant, ongoing and valuable support and motivation. I would also like to thank Rob Bellemann for his academic guidance throughout this project.

REFERENCES