D3.3 INTERIM BODY ANIMATION VISUAL SYNTHESIS DEMONSTRATION

Grant Agreement nr	856879
Project acronym	PRESENT
Project start date (duration)	January 1st 2018 (36 months)
Document due:	31/05/2021
Actual delivery date	31/05/2021
Leader	Cubic Motion
Reply to	steve.caulkin@cubicmotion.com
Document status	Submission version

Project funded by H2020 from the European Commission

Project ref. no.	856879
Project acronym	PRESENT
Project full title	**P**hotoreal **RE**altime **S**entient **ENT**ity
Document name	Interim Body Animation Visual Synthesis Demonstration
Security (distribution level)	Public (PU)
Contractual date of delivery	31/5/2021
Actual date of delivery	31/05/2021
Deliverable name	D3.3 Interim Body Animation Visual Synthesis Demonstration
Type	Demonstration
Status & version	Submission Version
Number of pages	12
WP / Task responsible	Cubic Motion (CM)
Other contributors	-
Author(s)	Ahmed Al-Obaidi, Steve Caulkin, Emelia Fiell and Francisco Peñaranda
EC Project Officer	David GUEDJ - David.Guedj@ec.europa.eu
Abstract	This Deliverable describes the research developed to produce synthetic body animation on the virtual agent starting from the performance of the real actor. Our methods, progress and early results are documented.
Keywords	Body Animation, Digital Human
Sent to peer reviewer	Yes
Peer review completed	Yes

Circulated to partners	No
Read by partners	No
Mgt. Board approval	No

Document History

Version and date	Reason for Change
1.0 04-05-2021	Document created by Ahmed Al-Obaidi
1.1 24-05-2021	Version for internal review
1.2 26-05-2021	Amended version post peer review by Ahmed Al-Obaidi

Table of Contents

1 EXECUTIVE SUMMARY

The work in this deliverable has moved our virtual actor from being able to respond with predefined actions to now being able to synthetically create new body animation. Going forward this will allow the virtual actor to react to a much wider range of use cases without requiring dedicated capture sessions to directly feed their responses.

There have been initial shoots with our actor, Gareth, from which we have built a library of motion capture data representing a wide variety of behavioural characteristics and targeted gestures. Unfortunately, we have been unable to capture more data due to Covid-19 restrictions but have upcoming shoots planned. A lightweight body asset has been created using the UE MetaHuman framework. This was then combined with the Framestore head, topology and rig, giving us our full actor asset for this work. Using this asset and motion capture data we performed a successful body solve which has then been polished with scene editing.

Synthetic animation was then generated from a computed delta based on captured poses applied to existing joint parameters, enabling the interpolation and extrapolation of new sets of body motion. This was then integrated into the PRESENT reference implementation and an implementation dedicated to interactively animate the agent body was developed.

We are encouraged by the initial results and are planning further work once we have been able to expand our dataset. This deliverable demonstrates that major milestones relating to body animation for the PRESENT virtual agent have been achieved. The outcomes of this deliverable will be widely related to various aspects of the project and in particular, the proposed use cases where the body non-verbal behaviours will be essential features of the virtual agent.

2 BACKGROUND

D3.3 *Interim Body Animation Visual Synthesis Demonstration* is the third deliverable for WP3 *Agent Visual Creation* and related to WP3T4 *Behavioural learning and knowledge*. This deliverable aims to demonstrate the software component prototype to animate and puppeteer the body of a virtual agent. The scope of this deliverable includes the creation of the agent lightweight asset, the motion data acquisition and processing, the generation of synthetic body animation and finally, the integration of the developed work into PRESENT reference implementation.

The main purpose of this deliverable is to showcase the current progress and research activities to produce synthetic body animation on the virtual agent based on the performance of the real actor. The presented progress in this deliverable will be continued through M30 and concluded in D3.6 *Body Animation Visual Synthesis*, the final report on the software component to body animate and puppeteer the agent.

3 INTRODUCTION

This deliverable describes the three main ingredients developed and used to produce a plausible body animation for the virtual agent as covered in Section 4. The ingredients consist of the implemented processes to create a body animation dataset, a generative mechanism to synthesise body motion sequences and the foundation of the Unreal Engine component development. Additionally, the report is supported by three recorded demonstrations showcasing the major elements of the deliverable as listed in Section 7.

3.1 Main objectives and goals

The main objectives of this deliverable are as follows:

1. To establish a pipeline to create a body animation dataset for the virtual agent based on the motion capture performance.
2. To create a lightweight digital asset of the actor, Gareth Leighton, to demonstrate and improve the efficiency of the proposed pipeline.
3. To develop a software component prototype to synthesise novel motion sequences based on performance examples to animate and puppeteer the lightweight digital asset.
4. To finalise the groundwork to integrate the lightweight asset into Unreal Engine (UE) and implement the basic functionalities to utilise the generated animation.

4 BODY ANIMATION

This section covers the proposed steps to synthesise and demonstrate a plausible body animation for a virtual agent. The processes involve the creation of the digital character, the plan and the production of the motion capture data, the curation of the subsequent dataset and, finally, the algorithmic development of the proposed method for generative body animation. Moreover, the current state of the integration process into the reference implementation is reported.

4.1 Lightweight asset

The role of the lightweight body asset as proposed is to provide a parallel rig to the high-quality asset to be produced by Framestore. The lightweight body rig was produced using the newly released in-house tools of the UE MetaHuman framework. MetaHuman Framework is a new tool created by Epic Games that will empower the user to create a bespoke photorealistic digital human, fully rigged and complete with hair and clothing, in a matter of minutes. The switch to using MetaHuman tools enables the creation of a generic skeleton and a corresponding mesh using auto rigging methods which empower PRESENT with several advantages. The benefits of MetaHuman include greater character diversity, cross-platform adaptability, wider applications and use cases and the possibility of continuous expansion as the tool advances.

Figure 1. Lightweight asset rendered in Unreal Engine: Body (Left) and Face (Right)

The next step for lightweight asset development includes combining the facial and the body rigs, both at skeletal and topological levels. Moreover, a skin texture and some initial grooming, eyebrows and eyelashes, were incorporated based on Gareth's scans as supplied by Framestore. Finally, a set of basic clothing and a pair of shoes were added to the character to complete the visual appearance. Figure 1 shows the current progress for the overall lightweight asset as rendered in Unreal Engine 4.

4.2 Processing pipeline

It is crucial to acquire a large library of motion capture (mo-cap) data that represents a wide variety of behavioural characteristics in order to build a generative model for body animation. This defines the starting point of a multi-stage process to establish an animation dataset to be used as a knowledge base for this purpose. The design of such datasets will have a large influence on the range of the animation poses to be synthesised based on the acquired behavioural performance. A carefully selected list of non-verbal gestures was captured followed by a data processing phase which was performed based on the pipeline described in this section.

A Vicon system is deployed at Cubic Motion's studio and used to capture marker-based mo-cap data. The physical markers are tracked using the software package 'Shogun' developed by Vicon for this purpose. Moreover, a lightweight asset version of the body rig is created and used jointly with the mo-cap dataset as input to the body processing pipeline. The implemented pipeline to create body animation based on motion capture techniques consists of four main stages as illustrated in Figure 2 and described below.

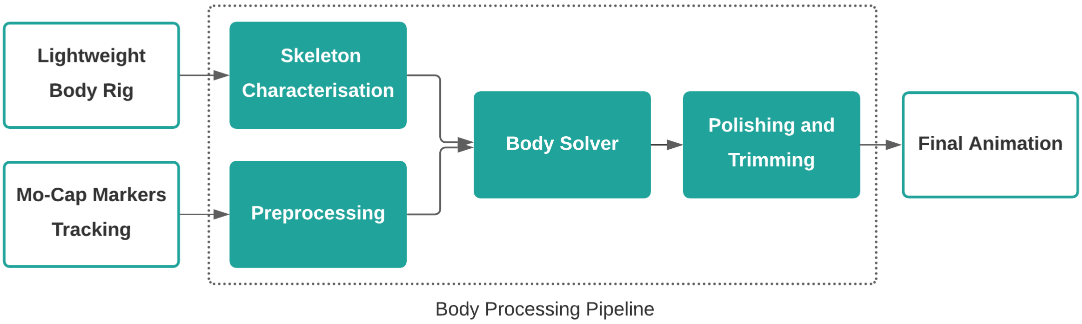

Figure 2. Body animation flowchart

1. **Skeleton Characterisation**: this step is responsible for modelling the body skeleton of the lightweight asset into human kinematics. The body characterisation step is conducted using Autodesk Motion Builder.

2. **Data Preprocessing**: an initial synchronisation step is required across the different data formats. This phase includes the extraction of the relevant frame numbers across various data modalities and the recovery of the absolute time code for the captured data.

3. **Body solver**: this step is performed to produce a skeletal animation using the mo-cap data onto the targeted character. It involves the process to animate the characterised 3D model by solving the optical motion capture data to a body skeleton to create articulated motion.

4. **Polishing and Trimming**: the final stage to produce body motion sequences includes the process to enable human intervention if necessary for further polished results. Furthermore, a scene editing task can be performed to mark and extract the targeted performance from the captured takes.

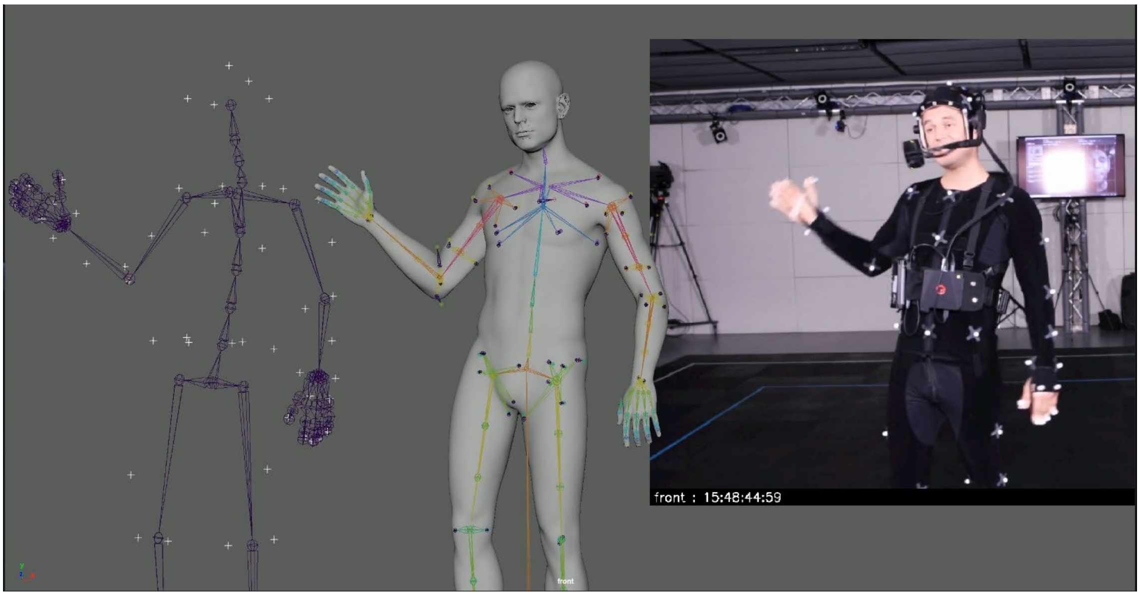

Figure 3. Body processing pipeline stages: mo-cap tracked markers (left), solved body animation (centre) and reference camera view (right)

A snapshot of the character motion evolution as a result of the implemented pipeline is illustrated in Figure 3. The leftmost illustration represents the mo-cap markers driving the baseline skeleton with the corresponding camera view on the right. The final animation solved on the body of the target character is shown in the centre of Figure 3. A video demonstration of the described body processing pipeline is provided in Section 7.

4.3 Body animation dataset

A plan was drawn to systematically capture motion data and eventually create a comprehensive body animation dataset. Initially, a study was conducted to review the various body gestures modalities from multidisciplinary perspectives including psychology and anatomy. A mind map representing a potential taxonomy for the various body modalities was produced as shown in Figure 4. The first layer of the tree represents the main body elements (i.e. head, torso and limbs), while the second layer extends further granularity and specificity levels of the body elements.

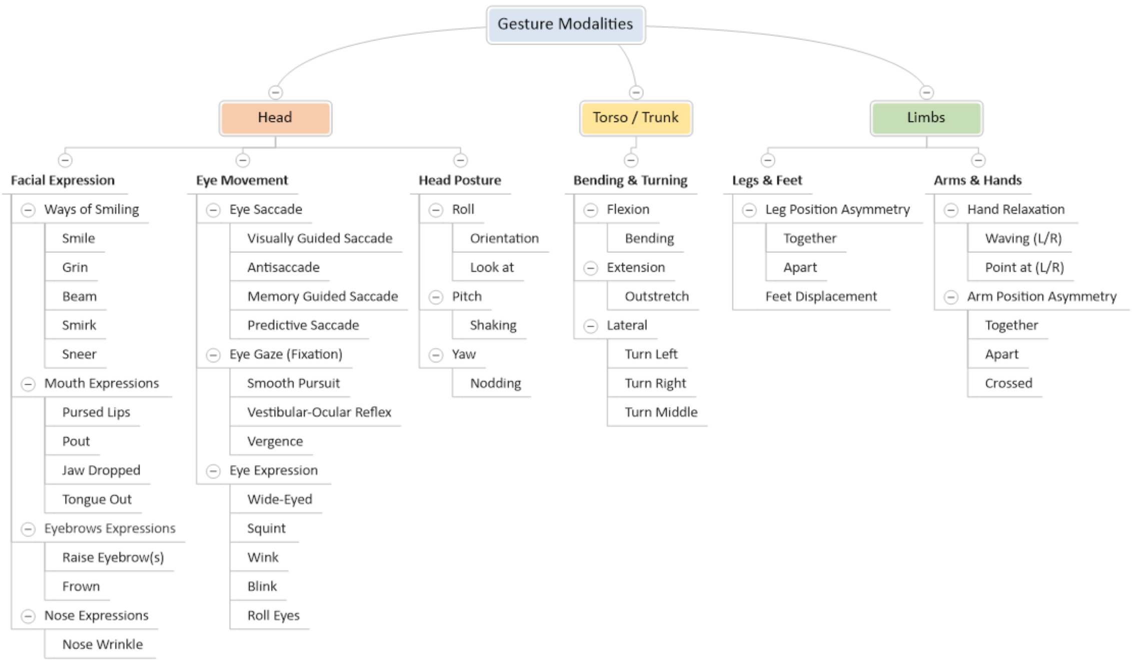

Figure 4. Gesture modalities taxonomy

As a result, a list of targeted gestures is compiled and used to direct the data capturing exercises. Classification of various non-verbal behaviours relating to head and body gestures is proposed in Table 1.

Table 1. Non-verbal behaviour gestures

Modality	Idling (loopable)	Moving (actionable)
Head	• Looking around • Relaxed	• Listening to user • Looking at object • Nodding • Shaking
Body (Torso + Limbs)	• Standing • Stretching shoulders and arms • Moving torso • Stretching legs • Sitting	• Lean towards user • Turn left / right • Point at - right / left arm • Hands together • Hands apart • Crossed arms • Wave - right / left arm • Sitting down • Standing up • Jump

Due to the COVID-19 restrictions, only one capture session involving body mo-cap has been carried out with the actor Gareth so far. Nevertheless, representative examples of all the gestures listed in Table 1 were captured. As a result, a dataset consists of a total of (150) non-verbal body and head gestures was created and used in the results of this deliverable. Figure 5 shows some examples of the dataset contents which will be used as a guideline for the upcoming capture sessions.

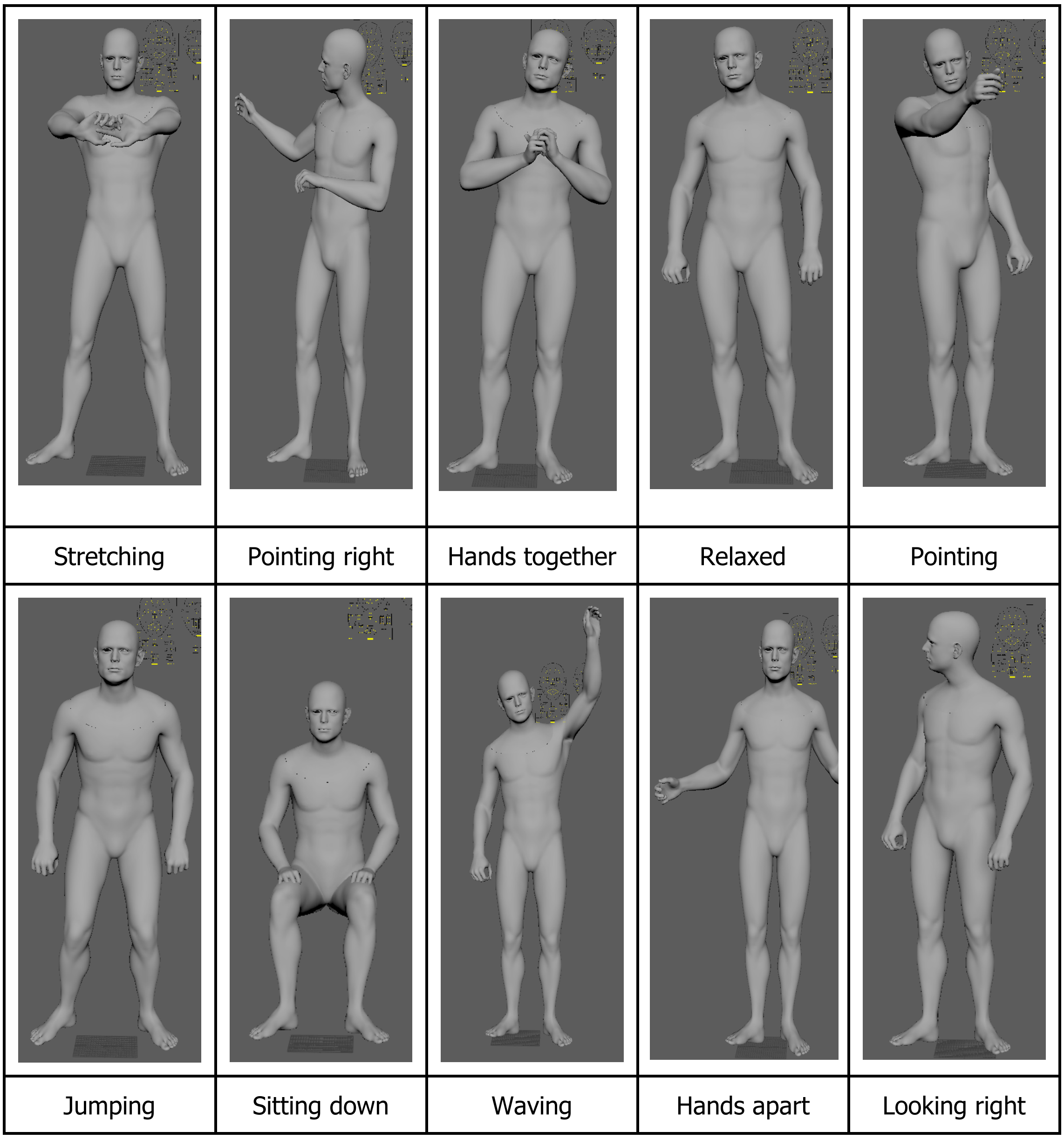

Figure 5. Nov-verbal gestures animation dataset

4.4 Generative body animation

Using the created gestures dataset, an attempt was made to develop an algorithm to synthesise new variations of body animation based on existing motion capture data. This experiment aims to generate new animation sequences in order to test and integrate the output into UE. Therefore, this experiment is considered a preliminary test for the development of the motion generation component as a whole and a crucial foundation for the next research endeavours.

An initial step was performed to analyse various non-verbal gestures within the dataset to find pairs of similar motion sequences in terms of motion direction and magnitude. A delta function was computed for a pair of sequences for each of the targeted skeleton joints. The computed delta is then applied to the existing joint parameters to enable the interpolation and extrapolation of new sets of body motion. Figure 6 shows an example output of the described algorithm where the middle character depicts an original mo-cap sequence while the characters on both sides show various degrees of synthetically generated variations. In addition, a video demonstration is included in Section 7.

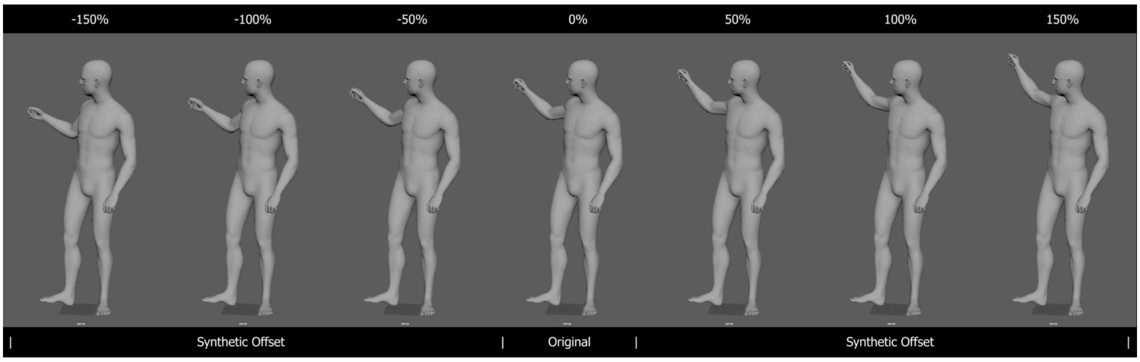

Figure 6. Generative synthetic animation

This experiment demonstrates the possibility to manipulate motion capture data by creating new gestures using a simple algorithm to blend in joint offsets over an animated sequence. Using a small training set, a new gesture is synthesised to aim at a different target point in space specified by preceding components within the system. This target point would be the input to the motion generation component, which would then generate the appropriate animation.

The use of gesture-based analysis represented in constraint space would provide a higher-level semantic parameterisation of the body movement rather than a lower level geometric space defining the position and orientation of the joints making up the skeleton. This experimentation besides the compiled body animation dataset has formed the basis of the next steps to build a generative model of body motion with the anticipation of larger database availability. Subsequently, machine learning techniques will be used to train a statistical model capable of generating body animation from higher-level emotional cues. This sits alongside the corresponding generative model of facial animation to provide all of the data needed to animate the virtual agent.

4.5 Reference implementation integration

The process of integrating Cubic Motion contribution into the PRESENT reference implementation includes multiple phases. Firstly, the lightweight asset as described in Section 4.1 was imported into Unreal Engine. In order to enable UE MetaHuman functionalities, the lightweight asset has been upgraded to a control rig with FK and IK capabilities for body, face, clothing and physics. Secondly, the solved optical motion capture data sequences have been imported as sequences into UE to puppeteer the agent asset. Finally, an implementation dedicated to interactively animate the agent body was developed.

Figure 7 shows a snapshot of the recently developed interactive test interface as included in Section 7. This interface allows testing the puppeteering of a particular gesture by selecting new target points and displaying the generated animation which represents a simulation of the motion generation component output. In this demo, four animation sequences have been imported into UE focusing on pointing behaviour to similar directions. A number of probes were implemented around the character to represent target objects in the form of golden and silver spheres. When one of the probes is clicked, the nearest neighbour animation will be played and when reaching the 'point at' pose, the pose will be corrected to point accurately to the target.

Figure 7. Interactive Unreal Engine demonstration

A set of 2D effectors is defined to determine the control values that affect corrections on horizontal and vertical axes. Then, a target map in the form of a grid is created defining which probe should be associated with a particular animation. These targets are the golden probes as they are placed as closely as possible to where the mapped animation is pointing at. On the other hand, if a silver target is clicked which is not mapped to a specific animation, the animation of the closest mapped target will be played and the gesture will be corrected accordingly.

As this is a work in progress, several improvements are currently being considered including further polishing of the animation sequence to improve the fluidity and the naturalness when correcting the target gesture. Moreover, a list of techniques will be investigated to simulate various body movements including learned motion matching for character navigation, head movement and gaze direction for engagement and eye contact, and lastly, the potential to incorporate the emotional cues into body gesture.

5 CONCLUSION

This deliverable demonstrates that major milestones relating to body animation for the PRESENT virtual agent have been achieved. The outcomes of this deliverable will be widely related to various aspects of the project and, in particular, the proposed use cases where the body non-verbal behaviours will be essential features of the virtual agent. Despite COVID-19 restrictions resulting in limited data acquisition, the planned research and development activities were successfully carried out as evidenced in this report. However, the lack of a more extensive dataset has hindered the possibility to explore more advanced machine learning techniques.

Based on the current progress, the next steps will focus on improving the lightweight asset visual appearance, including the addition of hair grooms. Following the recent regional updates regarding COVID-19 restrictions, a series of performance capturing sessions are planned to take place soon after this deliverable due date. This activity will mainly aim to extend the existing dataset to further develop a generative model for the body as well as facial animation. Furthermore, special attention will be directed to the implementation of the motion generation component which will be responsible for producing body and face animation using higher-level emotional cues. Finally, the overall performance will be validated by integrating more use cases to perform tests in offline as well as real-time settings.

6 ACRONYMS AND ABBREVIATIONS

Mo-Cap	Motion Capture
UE	Unreal Engine
FK	Forward Kinematics
IK	Inverse Kinematics

7 ANNEX

Links to the accompanying video demonstrations:

- Body processing pipeline: LINK
- Generative synthetic offset: LINK
- UE interactive demo: LINK