PREMIERE

Performing arts in a new era

Human Body Pose & Sentiment Analysis

6 October 2023 – M12

Document identifier: PRMR-D3.2-Human Body Pose & Sentiment Analysis v1
Version: 1.0
Author: Andreas Aristidou (CNS)
Dissemination status: PU

Grant Agreement n°	101061303
Project acronym	PREMIERE
Project title	Performing arts in a new era: AI and XR tools for better understanding, preservation, enjoyment and accessibility
Funding Scheme	HORIZON-CL2-2021-HERITAGE-01 (HORIZON Research and Innovation Actions)
Project Duration	01/10/2022 – 30/09/2025 (36 months)
Coordinator	Athena Research Center (ARC)
Associated Beneficiaries	• Stichting Amsterdamse Hogeschool voor de Kunsten (AHK) • Forum Danca - Associacao Cultural (FDA) • Tempesta Media SL (TMP) • Cyens - Centre of Excellence (CNS) • Kallitechniki Etaireia Argo (ARG) • Medidata.Net - Sistemas de Informacao para Autarquias SA (MED) • Fitei Festival Internacional Teatro Expresao Iberica Crl (FIT) • Instituto Stocos (STO) • Universite Jean Monnet Saint-Etienne (UJM) • Associacao dos Amigos do Coliseu Doporto (COL) • Stichting International Choreographic Arts Centre (ICK)

Project no. 101061303
PREMIERE

Performing arts in a new era: AI and XR tools for better understanding, preservation, enjoyment and accessibility

HORIZON-CL2-2021-HERITAGE-01

Start date of project: 01/10/2022

Duration: 36 months

History Chart				
Issue	**Date**	**Changed page(s)**	**Cause of change**	**Implemented by**
0.1	14/09/2023	ALL	1st Draft	Andreas Aristidou
0.5	22/09/2023	ALL	Additions	Andreas Aristidou
0.8	26/09/2023	ALL	Additions/Revision	Merce Alvarez
1.0	30/09/2023	ALL	Final Version	Andreas Aristidou

Validation			
No.	**Action**	**Beneficiary**	**Date**
1	Prepared	Andreas Aristidou (CNS)	30/09/2023
2	Approved	Aggelos Gkiokas (ARC)	01/10/2023
3	Released	Aggelos Gkiokas (ARC)	02/10/2023

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table of Contents

List of figures

List of tables

No table of figures entries found.

Executive Summary

This deliverable provides an overview of Tasks 3.2 and 3.3 within our project. Our ability to effectively train deep networks and models relies heavily on a diverse dataset encompassing a wide range of 3D motions, spanning various motion types. Currently, there are several databases commonly used for training deep models, and we leverage them to enhance our research.It encompasses two primary objectives:

1. **Contextual Analysis of Human Motion:** This involves a comprehensive examination of human motion, with a specific focus on the principles of Laban Movement Analysis.
2. **Human Sentiment Analysis**: This entails the capacity to assess, modify, or amplify emotions embedded within sequences of motion.

One of our primary data sources is the AMASS database, which acts as a central repository, consolidating data from diverse origins into a cohesive framework. This dataset revolves around the SMPL body model, enabling the conversion of motion capture data into realistic 3D human meshes. Of particular significance is the DanceDB motion capture database, an offshoot of AMASS, tailored specifically for contemporary dance performances. It incorporates emotional nuances meticulously analyzed through Laban Movement Analysis (LMA). Additionally, we incorporate the AIST++ Dance Motion Dataset, derived from real dancers who synchronized their movements with music, further enriching our collection of 3D dance data. Notably, our research benefits from the availability of a state-of-the-art 24-camera optical motion capture system, the PhaseSpace Impulse X2E. This advanced technology allows us to create a unique dataset comprising 30 distinct professional ballet performances, captured at the highest quality.

Task 3.2 focuses on the contextual analysis of human motion, with a specific emphasis on Laban Movement Analysis. The goal is to extract nuanced elements that effectively capture the complexity and dynamics of human movement, particularly in dance and theatre performances. To achieve this, we delve into deep learning architectures, including 3D Convolutional Neural Networks (3DCNNs) and autoencoders such as Variational Autoencoders (VAEs). These architectures are instrumental in providing a qualitative description of the subtleties in human motion.

Conversely, Task 3.3 centers around human sentiment analysis, enabling the manipulation, modification, or exaggeration of emotions within motion sequences. Drawing inspiration from CLIP models, we develop multi-modal embedding spaces that facilitate the estimation of semantic similarity between motion and style. This enables us to stylize motions without direct reliance on training data. To enhance motion synthesis quality and diversity, we employ diffusion models, neural motion graphs, and Variational Autoencoders. Additionally, we introduce physics-guided constraints to ensure that our generated animations adhere to the laws of physics.

In summary, our comprehensive approach combines diverse datasets, cutting-edge deep learning architectures, and innovative techniques to advance our understanding and synthesis of human motion. This research has the potential to find applications in fields such as animation, motion analysis, and beyond.

Acronyms and abbreviations

Abbreviation	Description
AI	Artificial Intelligence
CLIP	Contrastive Language-Image Pre-training
DB	Database
LMA	Laban Movement Analysis
MDM	Motion Diffusion Model
RCM	Russell Circumplex Model
VAE	Variational Autoencoder
3DCNN	3D Convolutional Neural Network
IK	Inverse Kinematics

1. Introduction

This deliverable focuses on D3.2 - Human Body Pose & Sentiment Analysis, and encompasses two primary objectives:

1. **Contextual Analysis of Human Motion**: This involves a comprehensive examination of human motion, with a specific focus on the principles of Laban Movement Analysis.
2. **Human Sentiment Analysis**: This entails the capacity to assess, modify, or amplify emotions embedded within sequences of motion.

Achieving these goals involves the extraction of intricate components that effectively capture the dynamic nature of human movement. Additionally, inspired by CLIP models, we aim to develop multi-modal embedding spaces that will enable the estimation of semantic similarity between motion and style, allows us to stylize motions without the need for direct reliance on extensive training data.

Section 2 discusses our primary motion capture database, AMASS, which consolidates data from 15 motion capture datasets, offering standardized skeletal and surface mesh representations. We also highlight our utilization of DanceDB within AMASS and the AIST++ database for dance-related research.

Moving to Section 3, we delve into Task 3.2, where our goal is to develop a neural network for in-depth human motion analysis, rooted in Laban Movement Analysis (LMA) principles. We explore advanced deep learning architectures, like 3D Convolutional Neural Networks (3DCNNs) and Variational Autoencoders (VAEs), to capture nuanced human motion features. We introduce the concept of deep motion signatures for motion sequence representation, emphasizing invariance to time scale and order.

In Section 4, Task 3.3 is detailed, aiming to enhance sentiment analysis in motion sequences. We plan to create a space connecting motion and style, enabling emotion-based stylization in dynamic performances. Our approach involves diffusion models and neural motion graphs for generating diverse animations, even with limited data. Physics-guided constraints ensure realism, especially in interactive scenarios.

Finally, Section 5 concludes this deliverable.

2. Training and Testing Databases

To effectively train our deep networks and models, we rely on a diverse dataset encompassing a wide array of 3D motions. In our research, we have leveraged a diverse set of motion capture databases to facilitate our investigations. Firstly, we have extensively utilized the AMASS database, accessible at https://amass.is.tue.mpg.de/. This database serves as a comprehensive resource, amalgamating data from 15 different optical marker-based motion capture (mocap) datasets into a unified framework and parameterization. The essence of AMASS lies in its ability to transform mocap data into realistic 3D human meshes, meticulously represented through a rigged body model known as SMPL (Loper, Mahmood, Romero, Pons-Moll, & Black, 2015) This model, renowned and widely adopted within the research community, not only offers a standardized skeletal representation but also furnishes a fully rigged surface mesh. Importantly, AMASS is adaptable to arbitrary marker sets, enabling the recovery of soft-tissue dynamics and lifelike hand movements. Please note that, in our research, we will use SMPL as the main motion format of research; however, we will also provide converters and parsers to move to different motion formats that are currently the industry standards.

Additionally, AMASS houses our DanceDB motion capture database, meticulously transformed into the SMPL-X format. This invaluable resource showcases over 140 contemporary dance performances, each infused with expressive emotions. This dataset has been pivotal in previous research endeavors, where we conducted thorough examinations of the intricate interplay between motion and emotions using the well-established Laban Movement Analysis (LMA) system.

In addition to AMASS, we have harnessed the AIST++ database, which can be accessed at https://paperswithcode.com/dataset/aist. The AIST++ Dance Motion Dataset consists of 3D dance data, meticulously reconstructed from real dancers who synchronized their movements with music. The dataset is derived from the AIST Dance Video DB, based on multi-views acquisitions, and features a sophisticated pipeline that encompasses the estimation of camera parameters, 3D human keypoints, and 3D human dance motion sequences.

When it comes to data for theatrical performances, we currently lack a dedicated, specialized database. However, there are valuable resources available, such as the AMASS database among others, that offer a broad and diverse collection of locomotion sequences. These sequences come equipped with labels indicating their corresponding actions. By utilizing databases like AMASS, we can access an extensive array of movements, which can be effectively employed for training purposes in the context of theatre performances. This diverse dataset not only facilitates the training process but also contributes to the realism and authenticity of the theatrical motions generated by our computational models. Furthermore, our research efforts have recently included the capture of professional ballet movements, facilitated through collaboration with the National Hungarian Opera (see Figure 1). Employing our state-of-the-art motion capture system, comprised of a 24 cameras Phasespace Impulse X2E motion capture system with active LEDs, we've compiled a database featuring 30 distinct performances of specific ballet movements. This resource is intended to serve as a valuable training dataset for our deep learning networks, further enhancing the richness of our motion capture research endeavors.

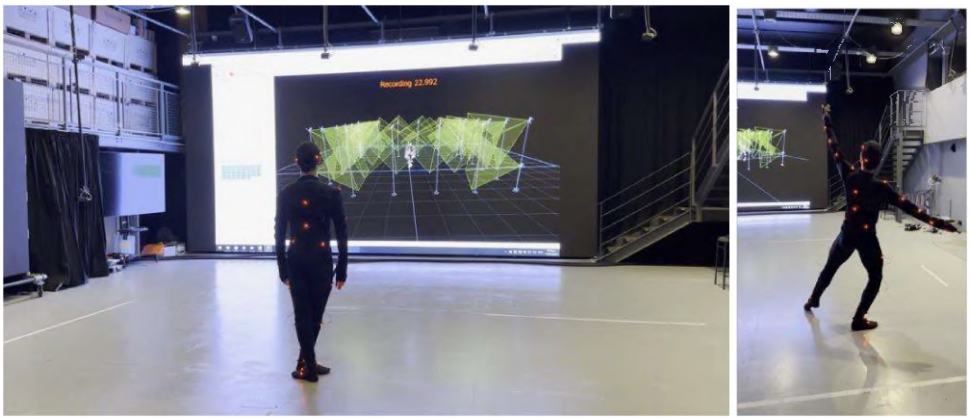

Figure 1: Dance motion capture in our recently establish virtual production and motion capture studio at CYENS

Lastly, we have introduced a recent innovation called SparsePoser (Pontón, Yun, Aristidou, Andújar, & Pelechano, 2023), a cutting-edge deep learning-driven method designed to reconstruct full-body poses with the use of just six tracking devices (see Figure 2). This system employs a convolutional autoencoder to generate precise human poses learned from motion capture data. Additionally, it incorporates a streamlined feed-forward neural network for inverse kinematics (IK), enabling precise adjustments of hands and feet in accordance with the data from the corresponding trackers. By employing this system, we can streamline the process of capturing a wider range of dance performances. It stands out for its cost-effectiveness, reduced manual labor requirements, and the minimized need for intricate sensors on the performer's body.

Figure 2: Highly accurate poses reconstructed from six 6-DoF trackers using SparsePoser.

3. Human body pose and motion analysis

This section pertains to Task 3.2, where our objective is to conceive and develop a neural network capable of qualitatively and contextually analyzing human motion. Our first target is to use the Laban Movement Analysis system to describe human movement nuances, using recent deep learning architectures. In addition, we are interested in developing a contextual analysis framework that will generate new motions of different durations, from examples, while maintaining invariance to timescale and temporal order. Our goal here is twofold: to create a system capable of generating motion sequences (e.g., dances or theater acts) of varying durations from a given input motion, while preserving the overall structure and content of the original motion. To achieve this, we plan to employ generative models, such as diffusion models, to acquire an understanding of the distribution of motion segments within a sequence. Subsequently, we will generate movements by selectively incorporating the essential and distinctive segments that best encapsulate and summarize the motion (whether it's dance or theater).

It's worth emphasizing that synthesizing long-term motion sequences using generative models like GANs or diffusion models is a formidable challenge. For instance, GANs often produce jittery motions and movements with temporal inconsistencies.

3.1. Emotion Recognition Technology using Artificial Intelligence (AI)

Human motion is inherently complex, and attempting to fully describe human actions using oversimplified motion descriptions is inadequate. Our systems must comprehend a wide spectrum of human actions, encompassing basic activities like walking, running, or jumping, as well as stylistic variations influenced by factors such as the performer's emotions, expressions, gender, and age. Moreover, they need to consider not only the fundamental aspects of motion (such as posture) but also the nuanced qualitative and quantitative characteristics that make each motion unique.

The proposed style-coherent motion analysis algorithm is designed to extract the qualitative attributes of a performance, guided by the principles of Laban Movement Analysis (LMA). Its objective is to identify the factors that constitute the distinct movement signature of the performer. LMA serves as a structured language for interpreting, describing, visualizing, and notating human movement. It offers a comprehensive framework for documenting human motion and is categorized into four main components: BODY, EFFORT, SHAPE, and SPACE. LMA strives to provide as simple a description as possible while capturing the necessary complexity of human movements. Here is an overview of each LMA component:

BODY Component: This component primarily focuses on the body and its connections in space. It describes the structural and physical characteristics of the human body, detailing which body parts are in motion, how they are connected, which parts influence others, the sequence of movements between body parts, and general statements about body organization.

EFFORT Component: The EFFORT component delves into the intention and dynamic qualities of movement. It explores aspects like texture, emotional tone, and how energy is employed in each motion. It encompasses four subcategories, each with two polarities, known as EFFORT factors:

- **Space**: This addresses the quality of active attention to the surroundings, with polarities of Direct (focused and specific) and Indirect (multi-focused and flexible attention).

- **Weight**: It relates to the perception of physical mass and its interaction with gravity. It is associated with the impact of movement and comprises two dimensions: Strong (bold and forceful) and Light (delicate and sensitive).
- **Time**: Time reflects the inner attitude of the body towards the temporal aspect of movement, not just its duration. Time polarities include Sudden (urgent, staccato, unexpected) and Sustained (stretching time, legato, leisurely).
- **Flow**: Flow concerns the continuity of movement, feelings associated with it, and progression. Flow dimensions encompass Bound (controlled, careful, and restrained movement) and Free (released, flowing, and fluid movement).

EFFORT changes often correlate with shifts in mood or emotion and are crucial for expressive movement.

SHAPE Component: SHAPE analyzes how the body changes its form during movement. It describes static shapes the body assumes, how the body changes in relation to itself and to points in space, and how the torso adapts to support movements in other parts of the body.

SPACE Component: SPACE describes movement in relation to the environment, pathways, and spatial tensions. Laban categorized movement orientation principles based on the body's kinesphere (the space within reach of the body, forming the mover's personal movement sphere) and the body's dynamosphere (the space where the body's actions occur, an essential aspect of personal style).

In this work, in contrast to previous works that use human designed features (Aristidou, Stavrakis, Charalambous, Chrysanthou, & Loizidou-Himona, 2015), (Aristidou, Charalambous, & Chrysanthou, Emotion analysis and classification: Understanding the performers' emotions using LMA entities, 2015) we will employ recent deep learning architectures to extract these LMA qualities, and better understand human movements.

3.2. Deep Learning Architectures

In our research, we will delve into the exploration of cutting-edge deep architectures designed to provide a qualitative analysis of human motion. Our approach involves training these networks on labeled data, where each label corresponds to specific elements derived from the LMA framework. The aim is to achieve a nuanced and precise description of human movement, capturing its intricacies and subtleties. The two key architectures we intend to investigate are:

1. **3D Convolutional Neural Networks (3DCNNs):** We are actively exploring the utilization of 3DCNNs due to their ability to capture spatial features in three-dimensional space. These networks have demonstrated their suitability for a wide range of tasks, including action recognition, pose estimation, and motion analysis. Leveraging their capacity to analyze motion across all three dimensions, we anticipate that 3DCNNs will enable us to extract rich and contextually relevant information from motion data.
2. **Autoencoders, Including Variational Autoencoders (VAEs):** Another avenue of investigation involves various autoencoder variants, with a particular focus on Variational Autoencoders (VAEs). These autoencoders aim to extract salient features from 3D motion data by learning compact representations. Past research has demonstrated that these compressed representations can serve as valuable resources for subsequent analysis or reconstruction tasks. By exploring the capabilities of autoencoders, we intend to uncover the latent features that contribute

to the essence of human motion, thus facilitating a deeper understanding and improved description of motion patterns.

Through the utilization of these advanced deep architectures, we seek to enhance our ability to qualitatively analyze and characterize human motion comprehensively. This research has the potential to significantly contribute to fields such as motion analysis, animation, and human-computer interaction by providing more accurate and insightful representations of human movement.

3.3. Contextual Analysis and Synthesis

In Task 3.2, our primary objective is to design and develop a neural network capable of contextually analyzing human motion, preserving its global structure and content while allowing for the generation of shorter or longer movements based on an input motion. This idea draws inspiration from the works of Shocher et al. (Shocher, Bagon, Isola, & Irani, 2019), Shaham et al. (Shaham, Dekel, & Michaeli, 2019), Kulikov et al. (Kulikov, Yadin, Kleiner, & Michaeli, 2023) in image processing, and has been previously investigated by Li et al.'s (Li, Aberman, Zhang, Hanocka, & Sorkine-Homung, 2022) GANimator and Tevet et al.'s (Tevet, et al., 2023) MDM in character animation.

Motion capture technology has proven to be a valuable tool for capturing dynamic movements, but the raw data lacks labels, annotations, or parameterization for further editing, synthesis, or control. A concise description of the performed movement is often necessary, both for indexing and retrieval purposes and for summarization. This is crucial for generating shorter clips with the same semantic content as the original animation or for extending a movement to meet specific duration requirements. Traditional motion analysis representations typically focus on skeletal geometry or pose matching, but these approaches often fail to capture the temporal evolution and stylistic nuances of motion.

To address these limitations, we recently introduced the concept of deep motion signatures (Aristidou, Cohen-Or, Hodgins, Chrysanthou, & Shamir, 2018), a succinct representation of motion sequences, and demonstrated in our recent work on dance synthesis (Aristidou, et al., 2023). This approach divides motion into a finite set of motion-motifs and defines signatures based on the distribution of these motifs (bag-of-motifs) within the sequence. Importantly, this method is time-scale and temporal order invariant, making it capable of handling variations in motion duration and speed, as well as actions occurring at random intervals.

Our goal is to design a network that generates shorter or longer motion sequences while preserving the distribution of motion-motifs from the input motion. This will enable us to create motion clips with the same semantic content as the original, a crucial task for various applications. We have already train fully convolutional GANs, similarly to the ideas of InGAN (Shocher, Bagon, Isola, & Irani, 2019) and SinGAN (Shaham, Dekel, & Michaeli, 2019), to learn the distribution of motion-motifs at different scales, allowing us to generate diverse motion samples while maintaining the overall structure and uniqueness of the training performance. We feed a generator with a motion sequence, utilizing convolution layers from the GANimator project to scale the animation to different durations. A discriminator assesses the distribution of motion patches in the input and generated motion. A decoder (inverted generator) then converges the scaled motion back to its initial size. Our loss function compares the original and synthesized motions to ensure high-quality results.

Currently, we are mostly interested in utilizing the idea of SinDDM (Kulikov, Yadin, Kleiner, & Michaeli, 2023) (see Figure 3) that uses diffusion models that have been proved to perform

better in motion synthesis rather than the GANs, e.g., the work of Tevet et al. (Tevet, et al., 2023). Additionally, we aim to ensure that the synthesized motion remains synchronized with the music beat, similarly to the works of Aristidou et al. (Aristidou, et al., 2023) and Zhou et al. (Zhou, et al., 2023).

In detail, our work aims at encompassing the design of a network to learn the distribution of motion segments within a sequence. It will generate movements by selecting and combining important and unique segments that best describe the motion. Furthermore, the network will be capable of extending the motion duration by adding common movements while preserving the semantic content of the input motion. We will conduct comparative analyses with existing methods such as the GANimator (Li, Aberman, Zhang, Hanocka, & Sorkine-Hornung, 2022) or MDM (Tevet, et al., 2023) networks and other methods on the pre deep-learning era e.g., seam carving (Avidan & Shamir, 2007).

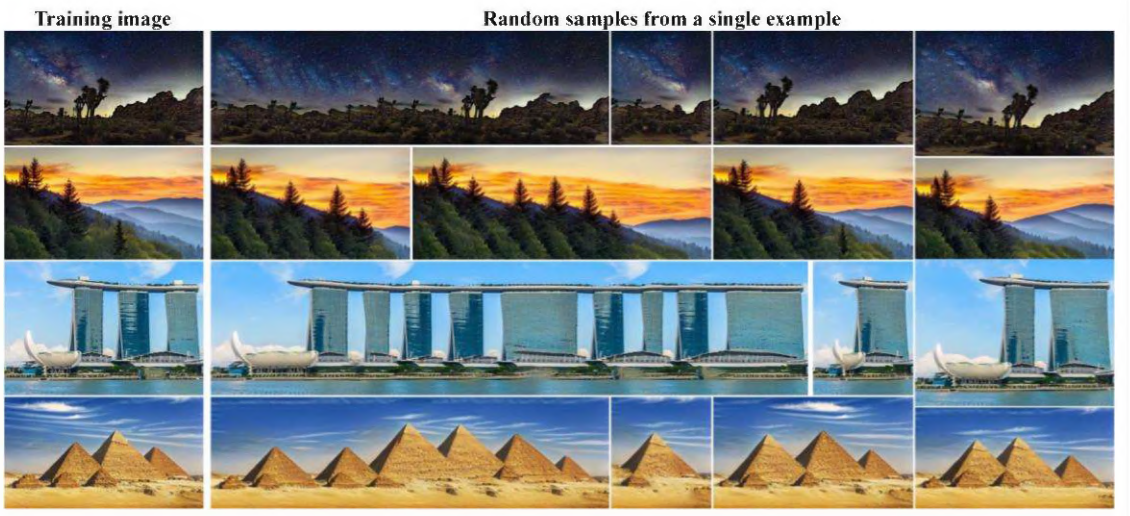

Figure 3: Single Image Denoising Diffusion Model

4. Human sentiment analysis

In Task 3.3, our overarching objective is to pave the way for advanced human sentiment analysis. Our aim is to enable the transfer, modification, or exaggeration of emotions within a given motion sequence, ultimately enhancing performance creation. Drawing inspiration from recent advancements in Contrastive Language-Image Pre-training (CLIP) models (see Figure 4), we aspire to cultivate a multi-modal embedding space bridging motion and style. This space will facilitate the estimation of semantic similarity between motions and emotions/styles, empowering us to stylize motions without direct reference to training data. This innovative approach decouples training data from input motions, eliminating the need for manual processing and motion registration. Furthermore, we intend to establish a bidirectional mapping between motion and emotions, employing the well-established Russell Circumplex Model (RCM) emotion coordinates (Aristidou, et al., 2017). This novel approach will enable the stylization of highly dynamic movements, such as those found in dance theatre, at interactive rates.

It is important to emphasize that our approach is inherently multimodal, involving a synergy of various sensory modalities. This collaborative effort will be closely coordinated with our esteemed partners, including ARC, specializing in audio and text modalities, and UJM, which specializes in emotion recognition from facial cues. Further insights into these collaborative endeavors will be elaborated upon in greater detail in the forthcoming deliverable, D4.1.

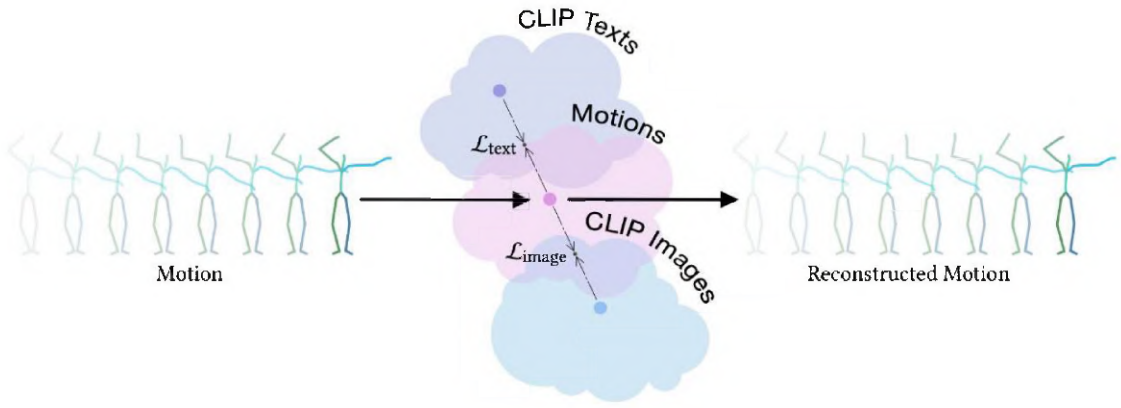

Figure 4: The CLIP space and model

To achieve these goals, we are exploring the potential of diffusion models and neural motion graphs. Diffusion models have proven effective in capturing motion segments within sequences. We will leverage these models to generate movements by extracting essential segments that best represent the input motion, whether it be dance or theatre. Additionally, inspired by CLIP models, our approach seeks to develop a multi-modal embedding space that measures semantic similarity between motions and specific tasks. This enables motion stylization for previously unseen movements without requiring reference to the training data, streamlining the process, and enhancing flexibility.

Our development includes a model that employs diffusion models to generate diverse motions while retaining the core motifs from a single learned input sequence. Given the limited data availability, we are working towards training the model on a single motion, similar to SinDDM (Kulikov, Yadin, Kleiner, & Michaeli, 2023). Our denoising network will be designed to cover only a portion of the input sequence's receptive field, allowing it to learn from multiple local temporal motion segments simultaneously. This approach will enable

temporal composition, style transfer for unseen styles, and the synthesis of variable-length motions.

Our animation synthesis approach will merge the strengths of diffusion models and neural motion graphs to address challenges related to generating high-quality, controllable, and diverse animations, particularly complex motions. We aim to employ denoising diffusion models to generate animations, even with limited annotated data, by creating sequences of prompted intervals and transitions. These sequences are then blended using neural motion graphs (Khan, Ribeiro, Kumar, & Francis, 2020), offering real-time and scalable motion synthesis capabilities. Each motion type will be represented as a separate neural node, thus reducing computational costs when incorporating new motion types. We will then introduce a single transition network to model transitions between motion nodes, and also to include a lightweight control module for fine-grained controllability. See for example Figure 5.

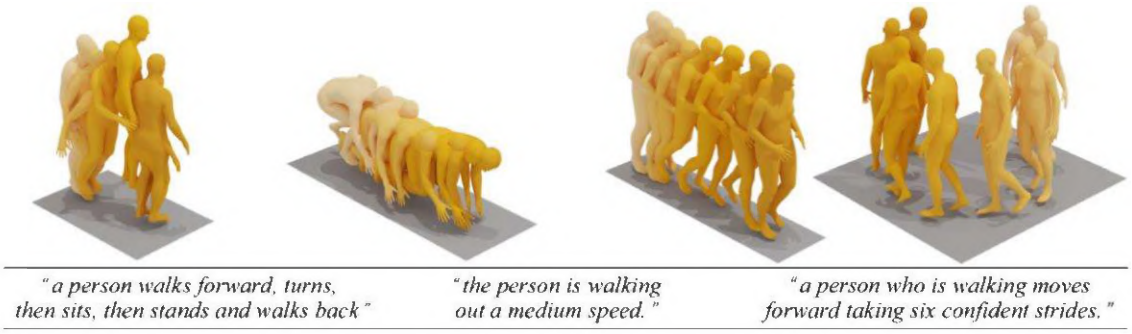

Figure 5: Executing text Commands via Motion Diffusion in Latent Space (Chen, et al., 2023)

Furthermore, we will enhance animation quality and diversity through latent space manipulation, by leveraging a Variational Autoencoder method similar to (Chen, et al., 2023). This method extracts informative latent codes that capture the essence of human motion sequences. By incorporating the diffusion process directly within the motion latent space, we expect to generate dynamic motion sequences that are aligned with conditional inputs, ensuring both realism and practical applicability.

Moreover, we will introduce physics-guided constraints by adopting a motion diffusion model that enforces physical constraints, like (Yuan, Song, Iqbal, Vahdat, & Kautz, 2023). A physics-based motion projection module is expected to project denoised motion into physically plausible motion, enhancing the naturalness and fluidity of animations, especially in scenarios involving interactions with objects.

It's important to note that our proposed models will be primarily trained on dance and theatre specific data.

5. Conclusions

In conclusion, this deliverable provides a comprehensive overview of our approach in D3.2 - Human Body Pose & Sentiment Analysis. We have discussed our primary data sources, with a significant emphasis on the AMASS database, which plays a central role in our research. Our work encompasses the development of neural networks for qualitative human motion analysis, rooted in Laban Movement Analysis principles, as detailed in Task 3.2. Additionally, Task 3.3 focuses on enhancing sentiment analysis in motion sequences, incorporating innovative techniques like CLIP models, diffusion models, and neural motion graphs. Throughout our research, we aim to push the boundaries of motion analysis and synthesis, particularly in the context of dance and theatre data. Section 5 marks the conclusion of this deliverable, setting the stage for our ongoing efforts in advancing human motion research.

Bibliography

Aristidou, A., Charalambous, P., & Chrysanthou, Y. (2015). Emotion analysis and classification: Understanding the performers' emotions using LMA entities. *Computer Graphics Forum, 34*(6), 262–276.

Aristidou, A., Cohen-Or, D., Hodgins, J. K., Chrysanthou, Y., & Shamir, A. (2018). Deep Motifs and Motion Signatures. *ACM Trans. Graph., 37*(6), Article 187.

Aristidou, A., Stavrakis, E., Charalambous, P., Chrysanthou, Y., & Loizidou-Himona, S. (2015). Folk Dance Evaluation Using Laban Movement Analysis. *ACM Journal on Comp. & Cultural Heritage, 8*(4), Article 20.

Aristidou, A., Yiannakidis, A., Aberman, K., Cohen-Or, D., Shamir, A., & Chrysanthou, Y. (2023). Rhythm is a Dancer: Music-Driven Motion Synthesis with Global Structure. *IEEE Transactions on Visualization and Computer Graphics, 29*(8).

Aristidou, A., Zeng, Q., Stavrakis, E., Yin, K., Cohen-Or, D., Chrysanthou, Y., & Chen, B. (2017). Emotion Control of Unstructured Dance Movements. *Proceedings of the ACM SIGGRAPH / Eurographics Symposium on Computer Animation.*

Avidan, S., & Shamir, A. (2007). Seam Carving for Content-Aware Image Resizing. *ACM Trans. Graph., 26*(3).

Chen, X., Jiang, B., Liu, W., Huang, Z., Fu, B., Chen, T., & Yu, G. (2023). Executing your Commands via Motion Diffusion in Latent Space. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.*

Khan, A., Ribeiro, A., Kumar, V., & Francis, A. G. (2020). Graph Neural Networks for Motion Planning. *https://arxiv.org/abs/2006.06248.*

Kulikov, V., Yadin, S., Kleiner, M., & Michaeli, T. (2023). SinDDM: A single image denoising diffusion model. *International Conference on Machine Learning*.

Li, P., Aberman, K., Zhang, Z., Hanocka, R., & Sorkine-Homung, O. (2022). GANimator: Neural Motion Synthesis from a Single Sequence. *ACM Transactions on Graphics, 41*(4), Article 138.

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics*, 248:1--248:16.

Pontón, J. L., Yun, H., Aristidou, A., Andújar, C., & Pelechano, N. (2023). SparsePoser: Real-time Full-body Motion Reconstruction from Sparse Data. *ACM Transaction on Graphics*.

Shaham, T. R., Dekel, T., & Michaeli, T. (2019). SinGAN: Learning a Generative Model from a Single Natural Image. *Proceedings of the the IEEE International Conference on Computer Vision (ICCV'19).* Seoul, Republic of Korea.

Shocher, A., Bagon, S., Isola, P., & Irani, M. (2019). InGAN: Capturing and Remapping the "DNA" of a Natural Image. *The IEEE International Conference on Computer Vision (ICCV).*

Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-or, D., & Bermano, A. H. (2023). Human Motion Diffusion Model. *The Eleventh International Conference on Learning Representations*.

Yuan, Y., Song, J., Iqbal, U., Vahdat, A., & Kautz, J. (2023). PhysDiff: Physics-Guided Human Motion Diffusion Model. *Proceedings of the IEEE International Conference on Computer Vision (ICCV).*

Zhou, Q., Li, M., Zeng, Q., Aristidou, A., Zhang, X., Chen, L., & Tu, C. (2023). Let's All Dance: Enhancing Amateur Dance Motions. *Computational Visual Media,* 9(3).