PERSONALITY TRANSFER IN HUMAN ANIMATION: COMPARING HANDCRAFTED AND DATA-DRIVEN APPROACHES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

By
Arçin Ülkü Ergüzen
September 2024

Personality Transfer in Human Animation: Comparing Handcrafted and Data-Driven Approaches
By Arçin Ülkü Ergüzen
September 2024

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Uğur Güdükbay(Advisor)

Hamdi Dibeklioğlu

Yusuf Sahillioğlu

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

PERSONALITY TRANSFER IN HUMAN ANIMATION: COMPARING HANDCRAFTED AND DATA-DRIVEN APPROACHES

Arçin Ülkü Ergüzen
M.S. in Computer Engineering
Advisor: Uğur Güdükbay
September 2024

The ability to perceive and alter personality traits in animation has significant implications for fields such as character animation and interactive media. Research and developments that use systematic tools or machine learning approaches show that personality can be perceived from different modalities such as audio, images, videos, and motions. Traditionally, handcrafted frameworks have been used to modulate motion and alter perceived personality traits. However, deep learning approaches also offer the potential for more nuanced and automated personality augmentation than handcrafted approaches. To address this evolving landscape, we compare the efficacy of handcrafted models with deep-learning models in altering perceived personality traits in animations. We examined various approaches for personality recognition, motion alteration, and motion generation. We developed two methods for modulating motions to alter OCEAN personality traits based on our findings. The first method is a handcrafted tool that modifies bone positions and rotations using Laban Movement Analysis (LMA) parameters. The second method involves a deep-learning model that separates motion content from personality traits. We could change the overall animation by altering the personality traits through this model. These models are evaluated through a three-part user study, revealing distinct strengths and limitations in both approaches.

Keywords: computer animation, Big Five Personality Traits, motion modulation, Laban Movement Analysis, deep learning.

ÖZET

İNSAN ANİMASYONUNDA KİŞİLİK AKTARIMI: EL YAPIMI VE VERİ ODAKLI YAKLAŞIMLARIN KARŞILAŞTIRILMASI

Arçin Ülkü Ergüzen
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Danışmanı: Uğur Güdükbay
Eylül 2024

Animasyondan kişilik özelliklerini algılama ve değiştirme becerisi, karakter animasyonu ve interaktif medya gibi alanlarda önemli çalışmalara önayak olmaktadır. Sistematik araçlar veya makine öğrenmesi yaklaşımları kullanan araştırmalar ve gelişmeler, kişiliğin ses, resim, video ve hareket gibi farklı modalitelerden algılanabileceğini ortaya koymaktadır. Hareket ayarlamak ve algılanan kişilik özelliklerini değiştirmek için geleneksel olarak el ile oluşturulmuş sistemler kullanılmaktadır. Ancak derin öğrenme yaklaşımları, daha ayrıntılı ve otomatikleştirilmiş kişilik geliştirme olanakları sunmaktadır. Bu gelişmekte olan konuya dikkat çekmek için el ile oluşturulmuş modellerle derin öğrenme modellerini, animasyonlarda algılanan kişilik özelliğini değiştirme verimliliği üzerinden karşılaştırılmıştır. Kişilik tanıma, hareket değişimi ve hareket oluşturma gibi çeşitli yaklaşımlar incelenmiştir. Bulgulara dayanarak OCEAN kişilik özelliklerini değiştirmek için hareketleri yönlendirirken kullanılabilecek iki yöntem geliştirilmiştir. İlk yöntem, Laban Hareket Analizi (LMA) parametrelerini kullanarak kemik pozisyonlarını ve rotasyonlarını değiştiren, el ile oluşturulmuş bir araçtır. İkinci yöntem, hareket içeriğini kişilik özelliklerinden ayıran bir derin öğrenme modelidir. Kişilik özellikleri bu modelle değiştirilerek animasyonun geneli değiştirilebilmiştir. Bu modeller üç bölümden oluşan bir kullanıcı çalışmasıyla değerlendirilmiş, iki yaklaşımın da güçlü yanları ve kısıtlamaları ortaya konmuştur.

Anahtar sözcükler: bilgisayar animasyonu, Beş Büyük Kişilik Özelliği, hareket modulasyonu, Laban Hareket Analizi, derin öğrenme.

Acknowledgement

I want to thank Prof. Dr. Uğur Güdükbay for his guidance throughout my research.

I would also like to thank the members of the jury, Asst. Prof. Dr. Hamdi Dibeklioğlu and Prof. Dr. Yusuf Sahillioğlu, for their insightful comments and valuable questions.

I would also like to thank our research group, ModVis, for their support during my studies and research and for creating a collaborative and encouraging environment. I am grateful to my colleagues for their help, feedback, and companionship throughout this journey.

Lastly, I thank my family and friends for their constant support, patience, and understanding. And a special thanks to my beloved girlfriend. Your belief in me has been a steady source of motivation.

The Ethical Committee of Bilkent University approves the user study of this work with decision No: İAEK_2024_08_08_02.

I acknowledge the support of the 5G and Beyond Joint Graduate Support Program during my master's studies.

This work is supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant No. 122E123.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Creating animated characters with distinct and recognizable personality traits is a significant challenge in animation, interactive media, and virtual agents. As characters become lifelike, there is a growing need for these agents to express unique personalities through their motions and behaviors. Personality traits are traditionally portrayed through body movements, gestures, and facial expressions. Being able to alter the personality of animations can enrich the user experience, making interactions more engaging and believable.

1.1 Context and Motivation

The ability to model and modulate the motion of animations has grown significantly, particularly with the rise of machine-learning techniques. Some of these methods ([1, 2, 3, 4]) enable modulation across various animation styles. However, no existing data-driven framework can account for the diverse variations in personality for motions. Previous works, such as [5] and [6], have used handcrafted methods to modify characters' perceived personalities. Techniques like Laban Movement Analysis (LMA) provide structured ways to manually adjust parameters, such as Effort and Shape Quality, to subtly influence how a

character's personality is perceived. Despite this, traditional methods are often labor-intensive and lack scalability.

With the advancements in deep learning, we have seen that the personality of humans can also be predicted from motion, as shown in [7, 8]. By modulating these predicted personalities, deep learning methods can also influence motion in animations and move beyond the constraints of handcrafted techniques. The motivation for this thesis lies in understanding the efficacy of these modern data-driven techniques relative to established handcrafted methods and exploring their potential to enhance personality transfer in animated characters.

1.2 Introducing the Approaches

This thesis examines two main approaches to transferring personality traits in animation. The first is a *handcrafted* method, which utilizes Laban Movement Analysis to modify animations based on predefined rules for movement. Adjusting LMA's Effort parameters allows the handcrafted approach to alter character motion systematically to reflect traits of the OCEAN personality model.

The second approach is *data-driven*, leveraging deep learning techniques to separate motion content from personality traits. It offers an automated way to transfer personality traits into character movements. The deep learning model decouples the base motion from the personality and modulates the animation through a learned latent space, offering the potential for more precise and scalable alterations.

1.3 Contributions

The contributions of this thesis are threefold:

- *Development of a handcrafted tool:* We present a motion modulation tool based on LMA parameters that allows animators to manually adjust animations to reflect different personality traits.
- *Implementation of a data-driven model:* A deep learning model that separates motion content from personality traits is introduced, enabling automatic and scalable personality transfer in animation.
- *Comprehensive evaluation:* We conduct three different user studies to compare the efficacy of the handcrafted and data-driven approaches, providing insights into the strengths and limitations of each method in altering perceived personality traits in animations.

1.4 Organization of the Thesis

Following the introduction, Chapter 2 summarizes the necessary background, including the personality model, movement analysis model, and related works in personality recognition and modulation in animation. Chapter 3 explains the methodology, covering the datasets in this field of research and implemented architectures and detailing the process for both handcrafted and data-driven personality transfer. Chapter 4 elaborates on the evaluation process, describing the design of each user study and analyzing the results of the motion modification techniques. Chapter 5 concludes the thesis by reflecting on the findings, discussing the limitations of the current approaches, and suggesting possible directions for future research.

Chapter 2

Background and Related Works

2.1 OCEAN Personality Model

Researchers have developed various trait-based theories for human personality in the psychology literature. Some examples are Cattell Sixteen Personality Factor [9], Hans Eysenck's psychoticism, extraversion and neuroticism [10], Myers–Briggs Type Indicator [11], and the OCEAN Personality Model [12]. Among those, the OCEAN Personality Model has become the most commonly used model for personality recognition in computation.

The OCEAN personality model, in other words, the Big Five personality traits, describes personality under five broad dimensions, each representing a range between two extremes. These traits are Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N), the acronym for the word "OCEAN."

Each of the five personality traits is summarized in the following subsections. Additionally, details on how each trait influences body movement are provided, highlighting the connection between personality dimensions and physical expression.

2.1.1 Openness

Openness represents a person's intellectual features, such as creativity, curiosity, and imagination. A person with high openness may be adventurous, creative, and more diverse, whereas individuals with low openness are typically more conventional in their thinking and behavior. They prefer familiar routines and are less interested in exploring new ideas or experiences.

Individuals high in openness often display more fluid, varied, and creative movements. They might be more expressive, using broad gestures and open postures that convey curiosity and willingness to explore new ideas. Low openness might result in more rigid and repetitive movements. The body language can appear reserved or closed-off, with less willingness to engage in novel or spontaneous actions.

2.1.2 Conscientiousness

Conscientiousness indicates the carefulness and dependability of the individual. Conscientious people are more likely to be organized, self-disciplined, and reliable. On the other hand, people with negative conscientiousness are more likely to be clumsy, careless, and disordered.

High conscientiousness is often associated with precise, deliberate movements. These individuals may exhibit controlled gestures, maintaining a structured posture that reflects their discipline and reliability. Low conscientiousness might lead to more careless or haphazard movements. The body language could seem disorganized, lacking focus or attention to detail, reflecting a more spontaneous or chaotic approach.

2.1.3 Extraversion

Extraversion describes the social side and outgoingness of humans. People with positive extraversion, also known as extroverts, tend to be more outgoing, warm, energetic, and optimistic. They are often cheerful and take action rather than reflecting profoundly. They prefer being around others since they focus more on the outer world. Individuals with negative extraversion, called introverts, tend to be shy, quiet, and gloomy. They remain passive to outside elements and enjoy alone time. These make them less energetic and sad-looking from other people's perspectives.

Extraverted individuals often have energetic and expansive movements. They exhibit lively gestures, open postures, and a strong presence, showing enthusiasm and social engagement. Low extraversion can lead to more contained and subtle movements. The body language may be more restrained, with less emphasis on outward expressions, reflecting a preference for solitary or quiet environments.

2.1.4 Agreeableness

Agreeableness reflects the cooperation skills of individuals and the harmony between them. People with positive agreeableness are kind, warm, and cooperative. They can empathize more quickly and be more generous than the ones with negative agreeableness. People with negative agreeableness are also prone to be narcissistic and selfish. They lack empathy; thus, they tend to be more competitive.

High agreeableness tends to manifest in warm, cooperative movements. These individuals might display gentle gestures and relaxed postures, often mirroring others' movements to show empathy and harmony. Low agreeableness could lead to more assertive or even confrontational movements. The body language might be more rigid or defensive, with gestures emphasizing personal boundaries or a lack of concern for others' feelings.

2.1.5 Neuroticism

The only trait where a person's personality is affected negatively while the trait goes to the higher end of the extreme is neuroticism. It is based on the likelihood of negative emotions in people. Individuals with high neuroticism may experience anxiety, fear, loneliness, worry, envy, and similar dark emotions more frequently than the ones with low neuroticism. Individuals who have low neuroticism can sustain their emotional stability easily.

High neuroticism often correlates with tense and erratic movements. Individuals may exhibit nervous gestures, frequent shifts in posture, and an overall sense of unease or restlessness in their body language. Low neuroticism (emotional stability) is usually associated with calm and steady movements. The body language may be relaxed, with smooth and controlled gestures reflecting confidence and emotional balance.

2.2 Laban Movement Analysis

Laban Movement Analysis (LMA) is a theoretical model that aims to describe and help understand human movement. The method is used in various areas, such as dance, acting, sports, physical therapy, education, animation, and video games. LMA was first introduced by Rudolf Laban and had some basic categories. His students have changed the structure of these categories over the years. The finalized categories are *Body*, *Effort*, *Shape*, and *Space*. Each individual combines the rules under these categories to create the motions.

The *Body* category shows how humans can form and move their bodies. It is also responsible for whether the body has moved independently or is influenced by other parts to move. This category can show the limits of the performer.

The *Effort* displays the intention behind the movement under more psychological and emotional aspects. For example, petting a cat gently and scrubbing

a pan are relatively close to each other since similar body parts are used in both of these. However, these actions differ significantly in intention, control, and dynamics. The *Effort* category can be divided into four subcategories with two opposite polarities:

The *Space* subcategory reflects the focus and directionality of the movement. A movement can be either *direct* or *indirect*. Direct movements are straightforward, focused on one point, and explicit. In scrubbing a pan, the hand moves in a determined, focused path to ensure thorough cleaning. The action can be labeled as a direct move. On the other hand, indirect movements are flexible and can have multiple directions. While petting a cat, the hand moves with a gentle, flowing path, following the contours of the cat's body. Which means it is an indirect move.

The *Weight* subcategory refers to the force or pressure behind the movement. A movement can be either *light* or *heavy*. Light movements are gentle and delicate and use minimal force. The initial example, petting the cat, is a *light* movement. Opposite to light, heavy movements require more force and are more powerful. Scrubbing the pan can be an example of solid movement since it requires much force to clean it effectively.

The *Flow* subcategory concerns the continuity and fluidity of the movement. A movement can be either *bound* or *free*. If a movement is *bound*, it is controlled, contained, and restrained. The movement is bound in the scrubbing pan example, with a sense of purpose and resistance against the grime. The movements that are easy-going, unrestricted, and continuous are defined as free movements. Petting the cat can be described as free since it is relaxed and continuous, with a sense of ease and fluidity.

Finally, we have the *Time* subcategory. It is responsible for the speed or pace of the movement. A movement can be either *sudden* or *sustained*. Sustained movements are slow, drawn-out, and gradual, as in the petting example. There are also sudden movements, which are quick, abrupt, and immediate. The counter-example, scrubbing the pan, can be described as sudden because the movement

is quick and repetitive, focusing on removing the dirt efficiently.

The third category under LMA is *Shape*. It helps in understanding not only the static positions the body can take but also the dynamic processes of movement and transformation. It provides a nuanced way to observe and describe how the body interacts with itself and its surrounding space. The Shape can be broken down into several aspects:

Form: The static shapes that a body can take. *Wall-like*, *pin-like*, *ball-like*, and *spiral-like* are examples.

Modes of shape change: These modes describe how the body transforms its shape concerning itself or the environment. There exist three modes: *Shape flow*, *Directional*, and *Carving*.

Shape qualities: These are the characteristics of how the body changes shape. The movement can be called rising, sinking, spreading, enclosing, advancing, or retreating.

Space is the final category of LMA. It focuses on how the movement occupies, navigates, and interacts with the space. Like other categories, *Space* can also be divided into subcategories:

- *Kinesphere*: The personal space surrounding the body can be reached without any steps. If the movements are close to the body, they are called near-reach movements. If they somehow extend the body, they are called mid-reach movements. Furthermore, if they extend to the farthest limits of the body, then they are far-reaching movements.
- *Spatial intent*: It concerns the direction and the intention within the kinesphere. Based on this directional intention, a movement can either be central, peripheral, or transverse.

2.3 Personality Recognition

Personality recognition from visual and motion-based data has become an active area of research, with various approaches demonstrating how human personality traits can be inferred through advanced computational models. These methods range from analyzing static images like selfies to videos capturing facial expressions to more dynamic cues such as body motion. Researchers have developed sophisticated systems that can accurately predict personality traits by leveraging multi-modal data. This section explores recent advancements in personality recognition across different media and modalities, focusing on how innovative systems capture the subtleties of human behavior to reveal underlying personality characteristics.

2.3.1 Image-based Approaches

Guntuku et al. [13] explore the potential of using selfies to predict personality traits by examining basic visual features and intermediate cues. The researchers collected selfies and annotated them with emotional positivity, facial visibility, and facial expressions. They utilized low-level visual features, including color histograms, aesthetic principles (e.g., the rule of thirds, vanishing points), GIST descriptors, Local Binary Patterns (LBP), and Fisher Encodings of SIFT, SURF, and HOG descriptors to capture key elements of the images. These visual features were then used to identify mid-level cues that reflect traits from the Five-Factor Model of personality. This study highlights specific aspects of selfies, such as facial expressions and camera angles, which indicate personality traits. The research found that predictions of how others perceive an individual's personality were more accurate than self-assessments, likely due to the amplification of cues like emotional positivity.

Fu and Zhang [14] applied an enhanced Active Shape Model (ASM) combined with a Deep Belief Network (DBN) to identify personality traits based on facial

features. They achieved higher accuracy in detecting facial landmarks by improving the traditional ASM algorithm with Gabor wavelets and gradient features. After extracting facial features, the researchers trained a DBN model to classify four major personality traits: extraversion, openness, agreeableness, and conscientiousness. This hybrid ASM-DBN approach yielded highly accurate results, especially for agreeableness (90.63%) and conscientiousness (91.42%). The study highlights the strong connection between facial structures and personality traits.

2.3.2 Video-based Approaches

Suen et al. [15] examine the use of convolutional neural networks (CNNs) for automatic personality recognition (APR) in asynchronous video interviews (AVIs). Their system analyzes video recordings of real job applicants, extracting facial features through the TensorFlow AI engine. The model successfully predicts the Big Five personality traits based on facial expressions and nonverbal signals using self-reported personality scores.

Critical features for personality prediction include tracking facial expressions using 86 facial landmark points across video frames. The model is also pretrained in Inception-v3 and Dlib facial detection, focusing on grayscale images to minimize background distractions. This semi-supervised learning system achieves high accuracy (90.9-97.4%) without needing extensive labeled data.

Song et al. [16] introduce a novel method for automatic personality recognition based on the temporal evolution of facial expressions rather than static frames or short video clips, typically used in other models. Their approach adopts a self-supervised learning framework to capture unique, person-specific facial dynamics from videos. The system learns the temporal progression of facial actions through the Rank Loss function without requiring personality labels, which allows the model to learn from unlabelled video data.

After freezing a U-net-style network, it is trained to capture broad facial dynamics. This type of network also uses intermediate filters to extract individual-specific information. The learned weights are used for predicting Big Five personality traits, with experimental results showing that multi-scale facial dynamics provide richer information for personality prediction than single-scale dynamics. Combining data from different tasks also boosts prediction accuracy.

Salam et al. [17] explore how personality affects engagement in human-robot interactions using a fully automatic analysis system. The study involves two human participants interacting with a humanoid robot in a triadic setting. In these conversations, both human and robot personalities are automatically assessed. The system extracts nonverbal behavioral cues and predicts participants' Big Five personality traits. The model focuses on body movement, interpersonal distance, and attention exchanged between the participants and the robot.

Additionally, the research investigates group engagement, analyzing how the alignment (similarity or difference) of personalities between humans and robots impacts collective engagement in the interaction. The results show that extroverted robots and participants increase engagement, while introverted interactions lead to lower engagement. Incorporating personality predictions enhances engagement classification compared to using nonverbal features only.

2.3.3 Motion-based Approaches

Dotti et al. [8] propose a framework that integrates non-verbal behavioral cues with contextual information from video data. This model uses spatio-temporal motion descriptors to capture individual engagement, social group dynamics, and environmental interactions. The model predicts Big Five personality traits by encoding personal movement patterns and context-specific interactions. It was tested on various datasets and outperformed previous methods in predicting personality in social and non-social settings.

The research conducted by Erkoc et al. [7] focuses on personality recognition

using skeletal data and Laban Movement Analysis (LMA). Their approach relies on LMA effort features extracted from skeletal landmarks to analyze movement style. These features were input into Graph Convolutional Networks (GCNs) to predict Big Five Personality traits with a regression model. The study demonstrated that LMA-based skeletal data significantly improved personality prediction, achieving state-of-the-art results on the UDIVA dataset [18]. Their method also shows that personality can be accurately predicted using motion data alone, avoiding the potential privacy issues of image-based and audio-based approaches.

2.3.4 Multimodal Approaches

In recent years, multi-modal personality recognition systems have emerged, which combine data from various sources to improve prediction accuracy. Gürpınar et al. [19] developed a model that fused audio, scene, and face features to estimate first impressions based on personality traits. Building on this, Aslan et al. [20] enhanced the multi-modal approach with more advanced feature fusion techniques. They introduced a system to recognize apparent personality traits from videos by integrating various modalities, such as facial features, environmental context, audio, and transcripts. Their method employs modality-specific neural networks that independently extract features from each modality, later fused at the feature level for final predictions.

The model utilizes pre-trained CNNs, like ResNet and VGGish, to extract high-level spatial and audio features, while Long Short-Term Memory (LSTM) networks capture temporal dynamics. Training occurs in two stages. First, the modality-specific subnetworks are trained separately. Afterwards, the overall model is fine-tuned using these pre-trained subnetworks. When evaluated on the ChaLearn First Impressions V2 challenge dataset [21], this multi-modal approach achieved state-of-the-art performance predicting Big Five personality traits.

Shao et al. [22] propose a method for personality recognition by simulating person-specific cognitive processes using a graph-based neural network framework. Unlike traditional approaches that directly predict personality traits from

non-verbal behaviors like facial expressions or vocal tone, this method models the subject's cognitive process through a person-specific CNN. This CNN captures how an individual's personality influences their facial responses during interactions based on the non-verbal behaviors of their conversational partner.

The system creates a graph that preserves the unique parameters of the person-specific CNN and the geometrical relationships between its layers. This model is then used to recognize the true personality of the target individual, achieving better results than methods relying solely on automatic personality perception (APP) based on external non-verbal behaviors.

2.4 Incorporating Personality into Virtual Agents

In recent years, considerable research has been conducted on incorporating personality into virtual agents to enhance their believability and effectiveness in interactive environments. We discuss two prominent approaches in this domain.

Durupinar et al. [5] propose a perceptual framework that integrates the OCEAN personality model into the body movements of virtual agents. By utilizing Laban Movement Analysis (LMA), they systematically map personality traits to motion parameters, such as hand gestures and posture, creating agents with distinct movement styles that reflect specific personality characteristics. They built a small motion capture database using a 12-camera Vicon system to achieve this. The dataset includes nine actions: walking, knocking, pointing, throwing, waving, picking up a pillow, lifting a heavy object, pushing a heavy object, and punching. The collected data was cleaned with the assistance of a certified movement therapist. The cleaned motions were retargeted onto a neutral wooden figure to avoid bias stemming from body shape during personality perception. These motions were then adjusted using LMA-based motion parameters. Using these motions, they conducted a user study via Amazon Mechanical Turk (AMT) to map LMA parameters to the OCEAN personality traits. The authors employed the Ten-Item Personality Inventory (TIPI) [23] to measure perceived

personalities, where each question was rated on a 3-point Likert scale.

Their user study findings revealed a significant correlation between OCEAN traits and Laban Effort elements and between Laban Effort elements and motion parameters. To validate these results, the authors conducted an additional user study. In this study, they used three different actions, each modulated by distinct personality traits, and transferred these motions to three different characters. This variation in actions and characters ensured that the personality-driven animations remained recognizable across diverse scenarios beyond the specific actions and models initially used—the results of the second study closely aligned with their earlier findings.

Sonlu et al. [6] expand personality modulation beyond just movement. Their framework integrates personality traits across multiple modalities, including dialogue, voice, facial expressions, and body motion. Their framework uses handcrafted dialogues that fit each personality type. For example, an agreeable character speaks more politely and enthusiastically, while a neurotic character tends to show more hesitation and insecurity. The Watson Text-to-Speech API [24] generates the agent's voice, and vocal features like pitch and intensity are adjusted to match the character's personality. The framework applies a similar approach to PERFORM for the agent's movement, using LMA's Shape Quality and Effort parameters. They adopted PERFORM's mapping for OCEAN traits to Effort and developed their mapping for OCEAN traits to Shape Quality due to the lack of existing quantization. Facial animations are driven by emotions (such as happiness, sadness, and anger) and are aligned with the character's personality traits. The framework modulates facial expressions using shape keys to control muscle movements, adjusting emotion decay rates based on neuroticism levels. The effectiveness of these communication modalities in conveying personality was evaluated through a scenario-based user study using Amazon Mechanical Turk. The results indicated that personality was perceived most clearly when all modalities were used together.

2.5 Style Transfer Between Motions

There are no data-driven models that can transfer personality between motions. The methods described below provide valuable insights into how styles can be transferred between different motion sequences. Since personality traits are essentially an abstract and high-level style form, they can be encoded and transferred between motions.

2.5.1 Real-time Style Transfer for Unlabeled Heterogeneous Human Motion

Xia et al. [1] propose an approach for real-time style transfer for human motion using unlabeled, heterogeneous motion data. Their method introduces an online learning algorithm that constructs local mixtures of autoregressive (MAR) models to capture the spatial-temporal relationships between different motion styles. Unlike prior methods that rely on global linear models or require labeled data, this approach allows flexibility in dealing with complex, unlabeled motion data like transitions between walking, running, and jumping. By constructing local models based on the closest examples in the database, the system translates input poses into different styles with linear transformations. A local regression model predicts pose timings, adjusting the motion's speed based on the style.

Their system performs well across various movements such as walking, running, and punching and supports real-time control of stylistic output. The method outperforms alternatives like Linear Time-Invariant (LTI) models and Gaussian Process (GP) models in comparative experiments, particularly for heterogeneous motion data. Furthermore, their system allows for style interpolation, where users can blend multiple styles during runtime.

2.5.2 Unpaired Motion Style Transfer from Video to Animation

Aberman et al. [2] tackle the same problem. Traditionally, such tasks required paired data (motions with the same content in different styles), which limited the applicability of these models to the styles seen during training. The authors propose an unpaired framework that learns from a collection of motions with style labels, enabling the transfer of styles not observed during training.

The key innovation here is the disentanglement of motion into content and style latent codes. While the content code is responsible for preserving the structure of the motion, the style code modifies its deep features using temporally invariant adaptive instance normalization (AdaIN), which is known from image style transfer tasks such as StyleGAN [25]. Their model also allows style extraction directly from video, bypassing the need for 3D reconstruction.

2.5.3 GANimator: Neural Motion Style Transfer Using Generative Adversarial Networks

The GANimator [3] is a Generative Adversarial Network (GAN)-based model for animation generation and style transfer. Building on the SinGAN framework [26], which learns internal relationships within a single image at multiple resolutions, GANimator adapts this structure to work with animations. The model operates by learning to recreate motion at different resolutions, where each level focuses on capturing either broad movements or finer details, depending on the resolution. Lower resolution stages in GANimator determine the overall motion, while higher resolution stages refine minor details.

Unlike traditional content-style disentanglement methods, GANimator uses adversarial training to map source motions to stylized versions, allowing the generator to produce realistic, stylized animations. A discriminator network ensures the output aligns with the desired style, while content loss maintains the core

structure of the motion.

The GANimator system can also be used for personality transfer, where each personality trait would require training a separate model, as the system is designed to work on single examples. At this point, examples representing the extremes of the personality spectrum (lowest and highest averages) can be used to express each personality factor. However, this also means GANimator can only serve as a generative model for some datasets, limiting its use in large-scale applications.

2.5.4 Motion Puzzle: Arbitrary Motion Style Transfer Using Pieces of Motion Data

Kim et al. proposed the Motion Puzzle approach [4], which introduces a novel approach to motion style transfer, allowing for the manipulation of individual body parts rather than applying a uniform style to the entire body. By enabling per-body-part control, the system significantly expands the range of stylized motions that can be created, as it can combine styles from different motions, like assembling a puzzle. The framework employs a two-step transfer process using adaptive instance normalization (AdaIN) and an attention network (BP-ATN), which effectively captures both global and time-varying local motion traits, improving the transfer of dynamic and subtle styles. Unlike previous methods, Motion Puzzle does not rely on style labeling or motion pairing, making it adaptable to publicly available motion datasets and capable of performing zero-shot style transfer. The framework's flexibility allows for integration into real-time motion generation systems, making it useful for applications like animation and character design while demonstrating substantial improvements over previous techniques in capturing complex motion details.

2.6 Neural Motion Fields for Kinematic Animation

In recent advancements in animation and motion generation, Neural Motion Fields (NeMF) [27] has emerged as a powerful approach to representing and synthesizing kinematic motions. NeMF offers a novel paradigm by modeling motion as a continuous function over time. This approach departs from traditional methods that treat the motion as a series of discrete frames or states. This continuous representation is achieved with a small Multilayer Perceptron (MLP) model, which learns to map temporal coordinates directly to motion poses.

The work is inspired by another work, Neural Radiance Fields (NeRF) [28], which is a pioneering technique synthesizing novel views of complex 3D scenes. By modeling a scene as a continuous volumetric function, NeRF uses neural networks to learn a mapping from 3D coordinates and to view directions to RGB colors and densities. This approach allows NeRF to render highly detailed and photorealistic images of a scene from arbitrary viewpoints, making it a groundbreaking method in computer vision and graphics. NeRF operates by sampling points along rays cast from a camera and uses a neural network to predict the color and density at each point. NeRF can generate novel views that are indistinguishable from real-world photographs by integrating these predictions.

Like NeRF models a scene as a continuous function over 3D space and viewing directions, NeMF models motion as a continuous function over time. In NeMF, the motion is represented as a function $f(t)$, where t is a temporal coordinate. This continuous representation allows the neural network to predict the motion state at any arbitrary time point, enabling smooth transitions and interpolation between poses. Unlike autoregressive models that rely on past states to predict future motions, NeMF directly infers motions from time alone, offering greater flexibility and efficiency in motion generation.

2.6.1 The NeMF Module

The basic NeMF module acts like a simple decoder. For a motion, the MLP module aims to generate the 6D joint rotations x_t^r for each joint in the skeleton and the root orientation r_t^o from a given time frame t. Instead of directly passing the time frame t to the model, they use a positional encoding function to transform t and then pass the encoded output to the MLP. The purpose of positional encoding is to embed positional information for each frame, ensuring that the model takes the sequence of the frames into account. With this approach, one single motion can be reconstructed:

$$f(t) \rightarrow (x_t^r, r_t^o). \tag{2.1}$$

The authors created a reconstruction loss function to optimize the function $f(t)$. The results of $f(t)$, 6D rotation matrices x_t^r and r_t^o, are converted into $[3x3]$ rotation matrices with a Gram-Schmidt-like process described in the work of Zhou et al. [29]. The new rotations, $\hat{R}_t$ and $\hat{R}_t^o$, are then used to calculate the geodesic distances to find the rotation and orientation losses as,

$$\mathcal{L}_{rot} = \sum_{t=1}^{T} \arccos\left(\frac{\mathrm{Tr}(R_t(\hat{R}_t)^{-1}) - 1}{2}\right), \tag{2.2}$$

$$\mathcal{L}_{ori} = \sum_{t=1}^{T} \arccos\left(\frac{\mathrm{Tr}(R_t^o(\hat{R}_t^o)^{-1}) - 1}{2}\right), \tag{2.3}$$

where Tr is the trace of the matrix and T is the total frame count. From the $\hat{R}_t$, they also calculate the local rotation matrix $\hat{R}_t^l$ and then pass it through the forward kinematics (FK) module, suggested by [30], to obtain the regularized joint positions $\hat{x}_t^p$. L_1 loss is used to find the position reconstruction:

$$\mathcal{L}_{pos} = \sum_{t=1}^{T} \|x_t^p - \hat{x}_t^p\|_1. \tag{2.4}$$

The total reconstruction loss $\mathcal{L}_{rec}$ is then calculated by the weighted sum of these losses where each λ indicates the weights of each loss:

$$\mathcal{L}_{rec} = \lambda_{pos}\mathcal{L}_{pos} + \lambda_{rot}\mathcal{L}_{rot} + \lambda_{ori}\mathcal{L}_{ori}. \tag{2.5}$$

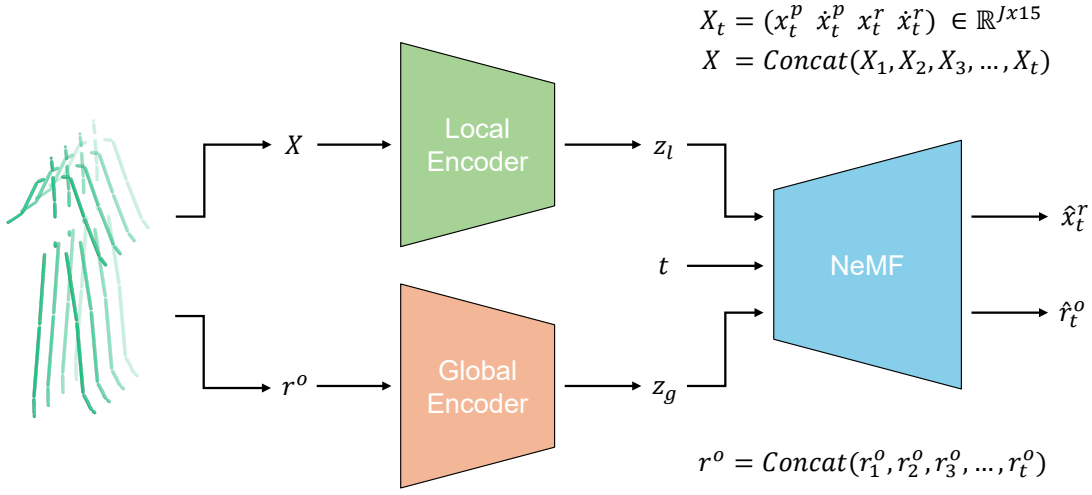

Figure 2.1: The overview of the generative NeMF model. The model takes animation sequences and encodes them in local (z_l) and global (z_g) latent variables. For each time frame t, the model learns to reconstruct joint rotations $\hat{x}_t^r$ and root orientation $\hat{r}_t^o$. Here, X_t is the local motion feature in frame t, whereas X is the concatenations of all. Similarly, r_t^o is the concatenation of all root orientations.

2.6.2 Generative NeMF

He et al. [27] show that NeMF can be used as a generative model to extend beyond a single motion sequence by incorporating a latent space that conditions the motion function on a latent variable z. The different values of z correspond to different motion styles or variations, enabling the model to produce a wide range of motion sequences. This generative process is framed within a Variational Autoencoder (VAE) [31] framework, where motion sequences are encoded into a

latent space and then decoded back to the motion domain, allowing for realistic and varied motion synthesis. The overview of the generative NeMF structure can be seen in Figure 2.1.

$$f(t, z) \rightarrow (x_t^r, r_t^o). \tag{2.6}$$

The latent z is constructed from two latent variables: local latent z_l and global latent z_g. The global and local latent variables are found using two separate convolutional encoders within the VAE framework. These encoders are specifically designed to disentangle the global and local aspects of the motion data:

Local Motion Encoding: The local encoder consists of specialized layers designed to process skeletal motion data. Essential layers include Skeleton Convolution and Skeleton Pooling [32] to capture spatial and temporal relationships within the skeletal structure. Skeleton Convolution focuses on local joint dependencies, while Skeleton Pooling reduces the complexity of the data by summarizing important features across joints.

The input to the local encoder is a sequence of local motion parameters, X, which is the concatenation of multiple time steps, X_t. Each X_t is constructed from joint positions x_t^p, joint velocities $\dot{x}_t^p$, 6D joint rotations x_t^r, and angular velocities $\dot{x}_t^r$, representing the pose of the skeleton relative to the root joint at each time step.

The outputs of the local encoder are the mean and variance that define a latent Gaussian distribution. The local motion latent variable z_l is sampled from this distribution, providing a compact representation of the underlying motion dynamics for further processing.

Global Motion Encoding: The global encoder is designed to process the overall trajectory of the motion, focusing on the root joint's orientation. The architecture combines 1-D convolutional layers to capture spatial dependencies and fully connected layers to integrate global motion information through time. These

layers work together to model the large-scale movements of the skeleton.

The input to the global encoder is the sequence of root orientations, r^o. The output of the global encoder is similar to that of the local encoder, producing the mean and variance that define a latent Gaussian distribution. From this distribution, the global motion latent variable z_g is sampled. The local latent variable z_l and global latent variable z_g are combined with the time frame t to reconstruct the motions.

The loss function in generative NeMF consists of two main components. The first loss component is the reconstruction loss, $\mathcal{L}_{rec}$, computed as described in the previous subsection. The second loss component is the Kullback-Leibler (KL) Divergence Loss, $\mathcal{L}_{KL}$. It ensures that the latent variables z_l and z_g follow the standard normal. $\mathcal{L}_{KL}$ is calculated as:

$$\begin{aligned} \log p(x_r, r_o) \geq \mathbb{E}_{q_{\theta_1}, q_{\theta_2}} \left[\log p(x_r, r_o \mid z_l, z_g)\right] \\ - D_{KL}(q_{\theta_1}(z_l \mid X) \parallel p(z_l)) \\ - D_{KL}(q_{\theta_2}(z_g \mid r_o) \parallel p(z_g)). \end{aligned} \tag{2.7}$$

Finally, the total loss is calculated as:

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda_{KL}\mathcal{L}_{KL}, \tag{2.8}$$

where λ_{KL} is the weight of $\mathcal{L}_{KL}$.

He et al. use the generative NeMF to perform motion interpolation and re-navigating tasks. Motion interpolation creates new motion sequences from examples by blending different styles or actions to produce smooth transitions. Motion re-navigating adjusts the trajectory to fit a new target or path while maintaining the original movement characteristics.

Chapter 3

The Methodology

3.1 Dataset Selection and Preparation

Transferring personality traits based on motion is complex because these traits are abstract and expressed through subtle cues like facial expressions, gestures, and movement style. Accurate modeling of these traits in animations demands datasets with detailed motion data and explicit personality trait labels. While several datasets are used in studying personality through motion, they often have limitations. Labeled datasets may need more detailed motion data for accurate personality transfer, while unlabeled datasets, rich in motion data, miss the necessary personality annotations. There is a need for a dataset that combines labeled personality traits with detailed motion data, particularly with robust skeleton representations. Such a dataset would improve the precision of personality-driven animation tasks, leading to more lifelike and personalized virtual characters.

3.2 Datasets

We can divide the existing datasets into two as *personality-labeled datasets* and *unlabeled datasets*.

3.2.1 Personality-labeled Datasets

Personality-labeled datasets are hard to find because collecting accurate personality information is challenging. There are two main ways to gather this data: self-reported assessments and perception-based evaluations. Self-reported assessments involve people rating their personality traits, usually through questionnaires, but the results can be affected by how self-aware and honest they are. Additionally, since each person is matched to a single personality profile, every action or behavior they exhibit will be labeled with that same personality. In other words, many participants must obtain diverse behaviors with different personality labels. Perception-based evaluations, where others assess someone's personality based on their behavior, appearance, or voice, introduce even more challenges. These evaluations can be biased by the observer's perceptions and the person's looks or voice. An expert in the OCEAN model would analyze these labels to reduce bias and improve accuracy. Because of these difficulties, personality-labeled datasets are not easy to create or maintain, making them rare. Several datasets use Big Five personality traits as labels. Each dataset differs in modalities, label collection type, and length. Table 3.1 summarizes the statistics of the datasets.

The *Synergetic sociAL Scene Analysis (SALSA)* dataset [33] captures 18 participants gathered in an indoor event. The participants freely interact with each other for over 60 minutes. The visual data is captured with four synchronized cameras. Each participant wore sociometric badges to capture audio and motions accurately. The dataset was then annotated based on positional data such as head and body orientations in 2D space and Big Five personality traits.

The biggest downside of this dataset is that it does not include 3D positional

Table 3.1: Labeled dataset statistics. Each dataset includes different modalities: Image (I), Audio (A), Sensory Information (SI), Video (V), and Skeleton (S). The labels of First Impressions are collected with Amazon Mechanical Turk (AMT), and for the rest, a questionnaire is used to obtain the self-reported labels.

Dataset	No. people	Time (hr)	Modalities	Label collection
SALSA [33]	18	90.5	I, A, SI	Self-reported
First Impressions v2 [21]	469	150	I, A	AMT
NoXi [34]	87	150	S, A, V, M	Self-reported
UDIVA [18]	149	18.3	S, I, A, M	Self-reported

data for participants. Estimating and extracting 3D data from 2D images can introduce inaccuracies and inconsistencies in the transferred personality traits between animated motions. These methods often rely on approximations that may not capture the full complexity of human motion, leading to potential distortions in the intended personality expression.

First Impressions v2 [21] comprises 10,000 video clips, each about 15 seconds long, featuring individuals speaking in English to a camera. These clips were sourced from over 3,000 high-definition YouTube videos and are split into training, validation, and test sets. The dataset is annotated with personality traits based on the Big Five model. These traits were labeled using Amazon Mechanical Turk (AMT). The results are processed using a procedure that results in reliable annotations.

Similar to SALSA, this dataset lacks 3D skeleton positions. Also, most videos miss full-body information, often capturing only partial upper-body movements, as they are predominantly vlogs. This limitation makes the dataset unsuitable for our needs since we work on full-body motion.

NOvice eXpert Interaction NoXi [34] provides a collection of multimodal recordings of dyadic novice-expert interactions, capturing both spoken language

and visual cues. Eighty-seven people were recorded during one-to-one interactions. The experiments were conducted in three countries spoken in seven different languages. The interactions were recorded with Kinect 2.0, and audio, video, depth, skeleton, and face information were obtained. In each session, participants provided demographic information and self-assessed their personality traits. These traits were evaluated using the Big Five model and measured with Saucier's Mini-Markers set of adjectives.

While NoXi offers a valuable dataset, its focus on natural, everyday interactions would only partially align with the specific requirements of our personality-driven animation task. We aimed to create highly expressive and stylized virtual characters, which demand more exaggerated and stylized movements than those typically captured in NoXi.

UDIVA [18] includes 188 sessions of dyadic interactions involving 147 participants aged 4 to 84. These sessions cover tasks such as talking, building legos together, playing a game about cards, and guessing animals. The dataset includes multimodal data such as video, audio, and physiological signals. It also provides detailed annotations, including facial, body, hand landmarks, and 3D eye gaze vectors. The personality traits were labeled based on self-reported questionnaires, with each trait recorded as a float value centered around zero. However, personality values for the validation and test sets were unavailable during participant interactions. Despite the advantage of having existing 3D motion data, the dataset's focus on seated interactions makes it less applicable to our specific problem requirements.

3.2.2 Unlabeled Datasets

The following datasets do not include any personality-based annotations. However, each dataset contains full-body motion captures. Labeling is required for such datasets to be used in our work. Table 3.2 provides the details of the unlabeled datasets.

Table 3.2: Unlabeled datasets statistics. The AMASS dataset is not labeled according to content and style. The ZeroEGGS (ZEGGS) dataset is only labeled according to style. SMPL-H [35] is a variant of SMPL.

Dataset	No. videos	Frames	FPS	Time (sec)	Avg. time (sec)	Content	Style	Data Type	No. joints
Xia [1]	-	79,829	120	665	-	28	8	BVH	31
BFA [2]	33	696,117	120	5,801	175.78	9	16	BVH	31
AMASS [36]	11,265	11 mil.	Varies	145,251	12.89	-	-	SMPL-H	52
ZEGGS [37]	67	484,740	60	8,079	120.58	-	19	BVH	75
Bandai-1 [38]	175	36,665	30	1,222	6.98	17	15	BVH	21
Bandai-2 [38]	2,902	384,931	30	12,831	4.42	10	7	BVH	21

The Xia et al. [1] dataset developed for the motion style transfer network discussed in Section 2.5 was captured using a motion capture system with eighteen cameras. It includes actions such as walking, kicking, and jumping, each performed in eight distinct styles, including neutral, depressed, old, and proud. All animations are annotated according to the action's content, style, and contact points. Similarly, the *BFA dataset* [2] was created for motion style transfer. This dataset can be seen as the expanded version of the Xia dataset. The dataset consists of 16 styles and nine content variations.

He et al. [27] use the *AMASS (Archive of Motion Capture as Surface Shapes)* [36] dataset to train a model for learning a latent representation of motion and various tasks. AMASS is a comprehensive collection of human motion data aggregated from 15 different motion capture datasets into a unified framework, providing detailed 3D meshes of the human body converted from raw motion capture data using the Skinned Multi-Person Linear (SMPL) model [39]. The dataset includes 344 subjects, 11265 motions, and approximately 40 hours of recordings.

ZeroEGGS [37] consists of 67 motion capture sequences of a female actor performing monologues in English, totaling 135 minutes. The dataset encompasses 19 distinct motion styles, covering variations in posture (e.g., Tired vs. Oration) and hand/head movements (e.g., Agreement, Oration). It was recorded at 60 frames per second (fps) using a 75-joint skeleton, capturing full-body motion, including detailed hand and finger movements. The dataset is also synchronized

with high-quality audio recordings for training and evaluating speech-driven gesture generation models.

The Bandai-Namco motion capture dataset [38] comprises two subsets: Dataset-1 and Dataset-2. Dataset-1, intended as a pilot dataset, contains 36,673 frames encompassing 17 diverse content categories (e.g., daily activities, fighting, dancing) and 15 motion styles. On the other hand, Dataset-2 focuses on providing a rich assortment of data for locomotion and hand actions, containing 384,931 frames across ten content types and seven styles. Dataset-2 provides data for each content-style combination, making it more suitable to train a motion-style transfer model.

Compared to established datasets such as Xia, BFA, Lafan [40], and 100STYLE [41], the Bandai dataset offers distinct advantages that make it suitable for our research. One of the key strengths of the Bandai dataset is its adherence to industry-standard human bone structures. This compliance ensures seamless integration with game engines and animation tools, eliminating the need for extensive retargeting processes often required with other datasets.

Additionally, Dataset-2 within the Bandai collection provides a well-balanced content-style distribution, making it an ideal choice for training high-quality style transfer models. This characteristic was critical in selecting this dataset for our research. While other datasets may offer greater volume or a more comprehensive range of styles, the Bandai dataset excels in practicality by directly addressing the challenges of applying motion style transfer in real-world applications.

3.2.3 Labeling the Bandai Dataset

We used the Bandai Namco dataset. We chose to annotate only Motion Dataset-1 because it was the pilot dataset with fewer samples, which made the annotation process faster. We also excluded specific motion samples that were too long, like dance sequences, or redundant, such as different walking directions. After these eliminations, we were left with 128 motions.

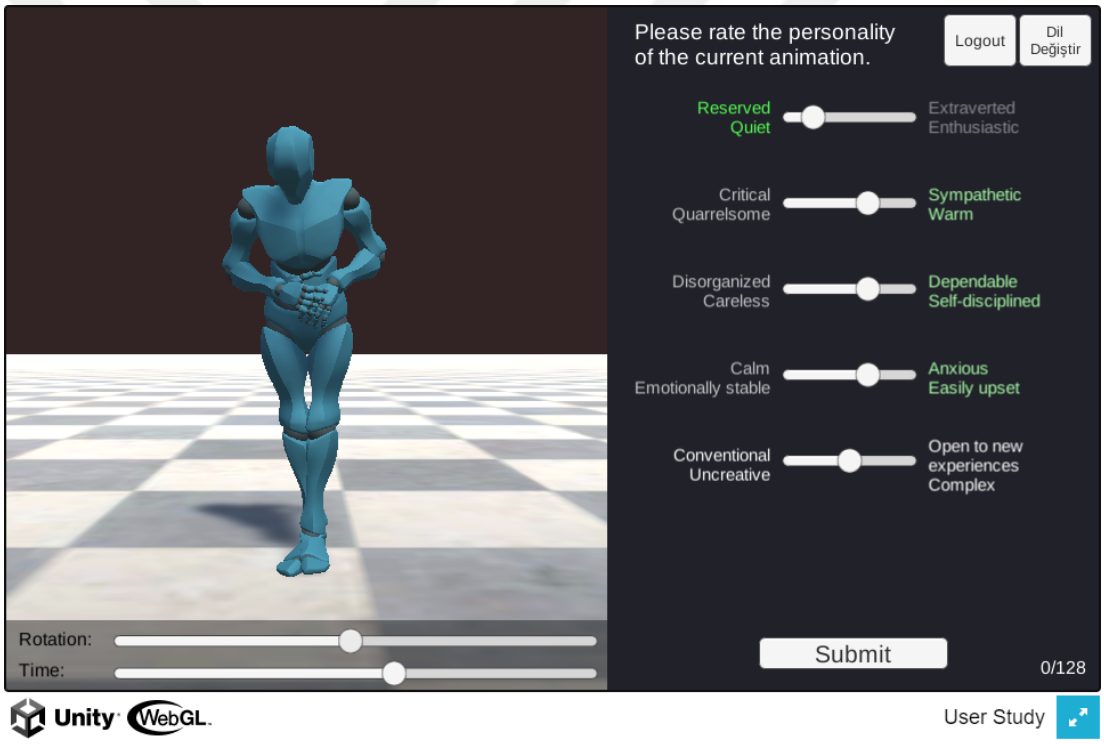

Figure 3.1: The website, built with Unity, is used to annotate the Bandai-1 dataset with Big Five personality traits.

To facilitate the annotation process, we developed a website using Unity (see Figure 3.1) and recruited participants through email invitations and Prolific, an online crowdsourcing platform. Prolific users were compensated based on their time rating 10 unlabeled animation samples, while participants recruited via email volunteered their time. We received 184 individual responses, resulting in 2,280 ratings. 17 to 21 participants evaluated each animation sample. Based on five questions corresponding to the Big Five personality traits, the ratings were recorded on a 7-point Likert scale. These responses are also used in [42].

The raw annotations had some outliers. While methods like RANSAC [43] or filtering by mean and standard deviation could be used to remove them, we opted for a more straightforward approach. Each personality trait label was initially scaled between $[-3, 3]$. For each trait, the dataset was divided into three regions.

$$R_1 = [-3, -1], \quad R_2 = [-1, 1], \quad \text{and} \quad R_3 = [1, 3].$$

Then we eliminate the region R_i if $|R_i| < 0.7 \times \max\{|R_1|, |R_2|, |R_3|\}$, where $|R_i|$ represents value count in the region R_i. Ideally, we aim to retain only one region, but in cases where no consensus is reached, two or all three regions are kept. After determining the regions, we calculate the sample average by averaging the regions. The initial 2,280 ratings were reduced to varying amounts for each personality trait: 1,798 (O), 1,831 (C), 1,753 (E), 1,739 (A), and 1,757 (N). Figure 3.2 shows the standard deviations for each personality trait before and after filtering the outliers.

3.3 LMA-based Handcrafted Motion Modulation Tool

The key objective of our motion modulation tool was to create variations of the same action performed in different personality styles. These modulated actions were crucial for comparing their effectiveness against those the deep learning

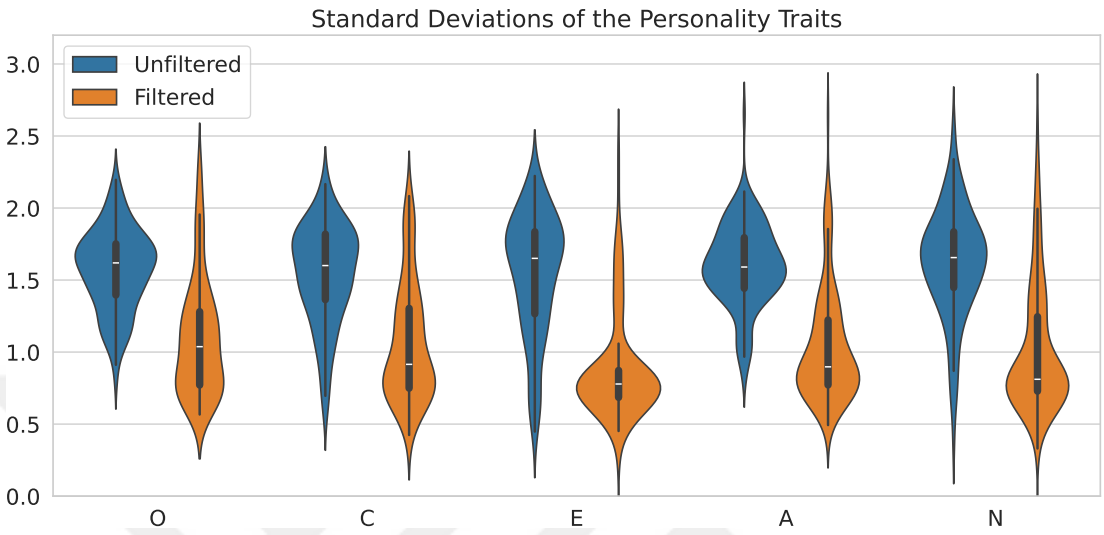

Figure 3.2: Violin plot showing the distribution of standard deviations for each personality trait (OCEAN) across animations in the Bandai dataset. The plot compares raw annotations (unfiltered) with annotations with outliers removed (filtered).

model generated. By the end of our work, this comparison allowed us to evaluate whether the augmented actions provided a more accurate or varied representation of personality-driven motion than the deep learning models alone.

We adopted the work of Sonlu et al. [6] to enhance our motion alternation approach. Their framework allows virtual characters to express the full range of the five personality factors through dialogue, voice, body motion, and facial expressions. Specifically, the body movement module in their system drives changes in the joints of the conversational agent, guided by two key Laban parameters: shape and effort. Our study focused on the Laban effort parameters as the foundation for the motion alternation tool. While the original work utilized the Unity Engine [44] to introduce variations and modify animations in real-time, we employed Blender [45], an open-source 3D software, to change the motions in our dataset, enabling us to implement these changes with a greater degree of control and flexibility.

The system starts by asking the user to select a base armature skeleton for the dataset that will be augmented. This selected armature becomes the main skeleton for the entire dataset. From this armature, a new skeleton is created.

The first skeleton controls bone movement using Forward Kinematics (FK), while the second uses Inverse Kinematics (IK). Eight additional bones are added to make IK work—four pole bones to control the direction of the knees and elbows and four target bones to control the position of the hands and feet. With these additions, the armature is ready for use (see Figure 3.3).

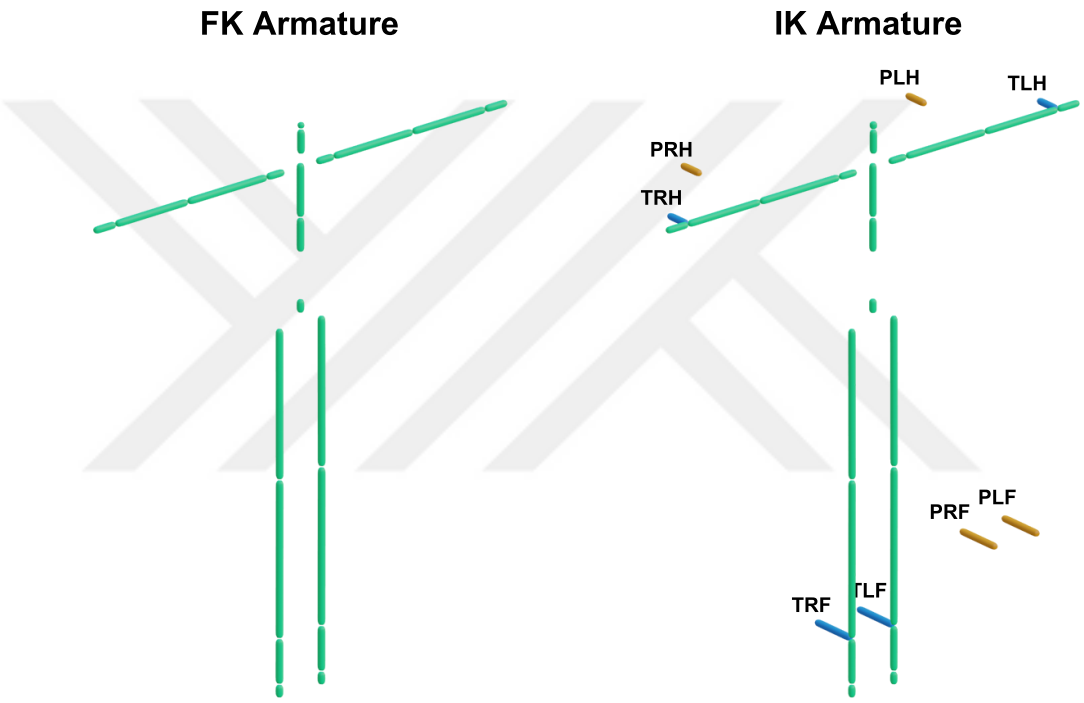

Figure 3.3: The Bandai dataset's armatures are visualized in Blender, showcasing both Forward Kinematics (FK) and Inverse Kinematics (IK). In the visualization, blue bones indicate the target bones, while orange bones represent the pole bones. 'R' and 'L' refer to right and left, respectively, and 'H' and 'F' denote hand and foot bones.

The animations in the dataset will be applied to the base armature and modified based on the influencing factor, either on the FK armature or the IK armature. There are a total of four factors that affect the animation. These are the *Space*, *Weight*, *Time*, and *Flow*. Each factor can take a value within the $[-1, 1]$ range. The influences of each factor are discussed in the following subsections.

3.3.1 *Space*

Being one of the Laban effort parameters, the *Space* indicates whether the motion is performed *directly* or *indirectly*. In the work of Sonlu et al. [6], the effect of the *Space* was achieved by spreading the hands on the horizontal axis. The target bones of the hands and feet in the IK armature system were made to move away from or closer to the body depending on the sign of *Space* factor f_S to achieve a similar change. If f_S is positive, the limbs spread, creating a more "indirect" appeal. When f_S is negative, the limbs will close, resulting in a more "direct" appeal.

First, each target bone inside the IK armature is paired with a corresponding "limit" bone, with the upper arm bones serving as limit bones for the hand targets and the upper leg bones for the foot targets. For each frame in the animation, a limit vector V_L is calculated by subtracting the position (head joint) of the limit bone from the position of the target bone:

$$V_L = J_{target} - J_{limit}. \tag{3.1}$$

To achieve spreading or enclosing motions, the limit vector V_L is decomposed into its vertical ${V_L}^v$ and horizontal ${V_L}^h$ components, which are then used to rotate the vector in the depth plane. The initial rotation degree r is set to 15°, representing the maximum possible rotation, and is further adjusted by multiplying it with the *Space* factor f_S.

Before determining the vertical and horizontal shifts for the target bones, we assess the distance d_{sym} between each pair of symmetric target bones. If the *Space* factor f_S is negative and the distance d_{sym} is less than the magnitude of V_L ($\|V_L\|$), a new symmetry factor f_{sym} is calculated.

$$f_{sym} = \begin{cases} \left| 1 - \frac{|\, d_{sym} - \|V_L\| \,|}{\|V_L\|} \right|, & \text{if bones are not crossing,} \\ 0, & \text{if bones are crossing.} \end{cases} \tag{3.2}$$

This factor prevents issues during movements such as walking, clapping, or bringing hands together. Specifically, when the horizontal distance between two symmetrical target bones is close to zero, the f_{sym} value will also approach zero.

The final rotation amount R is then obtained by multiplying the initial rotation r by the symmetry factor f_{sym}. The horizontal and vertical shifts for the target bones are calculated using this rotation amount as follows:

$$shift_h = ({V_L}^h \cdot \cos(R) - {V_L}^v \cdot \sin(R)) - {V_L}^h, \tag{3.3}$$

$$shift_v = ({V_L}^h \cdot \sin(R) + {V_L}^v \cdot \cos(R)) - {V_L}^v. \tag{3.4}$$

Finally, the new position of the target bone in that frame is determined by shifting its current position by the combined shift values: $shift_h$+$shift_v$. Figure 3.4 shows the effect of the *Space* factor in positive and negative directions.

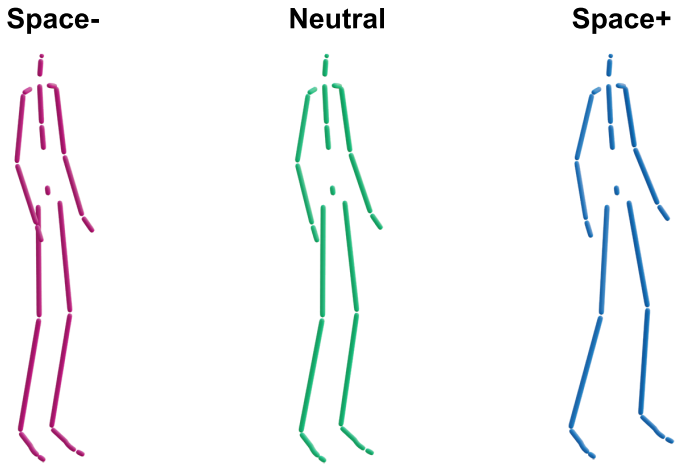

Figure 3.4: The impact of the *Space* factor: the left armature performs an enclosing motion due to the negative *Space* factor, while the right armature executes a spreading motion driven by the positive *Space* factor.

3.3.2 *Weight*

Another effort parameter, *Weight*, reflects the movement's interaction with gravity. The *Weight* parameter is categorized into *light* and *heavy*. The *Weight* factor f_W adjusts the IK armature system, altering the overall motion. A positive f_W value indicates a "descending" movement matched with "heavy", while a negative f_W value indicates an "ascending" movement matched with "light." During these adjustments, the foot target bones remain stationary. In contrast, the hip bone B_h and hand target bones move either lower or higher relative to their original positions.

We begin by adjusting the position of the hip bone, B_h. Directly altering the vertical position of B_h is not feasible since the foot targets must remain in contact with the ground. In "bowing" or "rising" movements, the hip reaches the maximum distance from the foot targets. To address this, we first determine the midpoint M of the action, calculated as the average distance between the two-foot targets. Next, we establish the limit distance d_{lim}, representing the distance between B_h and the floor.

The midpoint M determines the horizontal and depth-wise shift, $shift_{hd}$, for B_h. When the f_W gets closer to 1, B_h will also get close to M (Equation 3.5). Meanwhile, d_{lim} is used to calculate the vertical shift, $shift_v$. If f_W is negative, we also identify a limit bone for B_h, which will be either the left or right upper leg. The distance between the hip and this limit bone is d_{LimH}.

$$shift_{hd} = \begin{cases} (M - J_h) \cdot f_W, & \text{if } f_W > 0, \\ 0, & \text{if } f_W = 0 \end{cases} \tag{3.5}$$

$$shift_v = \begin{cases} \dfrac{d_{LimF}}{20} \cdot f_W, & \text{if } f_W \geq 0, \\ \max(d_{LimF} - d_{LimF}) \cdot f_W, & \text{otherwise} \end{cases} \tag{3.6}$$

The shift vector $shift_{all}$ is then calculated by simply combining the $shift_{hd}$ and

$shift_v$ vectors as components. The same vector $shift_{all}$ is also applied to hand target bones.

$$r = \begin{cases} 15^\circ \cdot f_W, & \text{if } f_W \geq 0, \\ 5^\circ \cdot f_W, & \text{otherwise} \end{cases} \tag{3.7}$$

In addition to these positional changes, the neck and spine bones are rotated with the angle of r along the body's horizontal axis where r is calculated as given in Equation 3.7. Figure 3.5 shows the effect of the *Weight* factor in positive and negative directions.

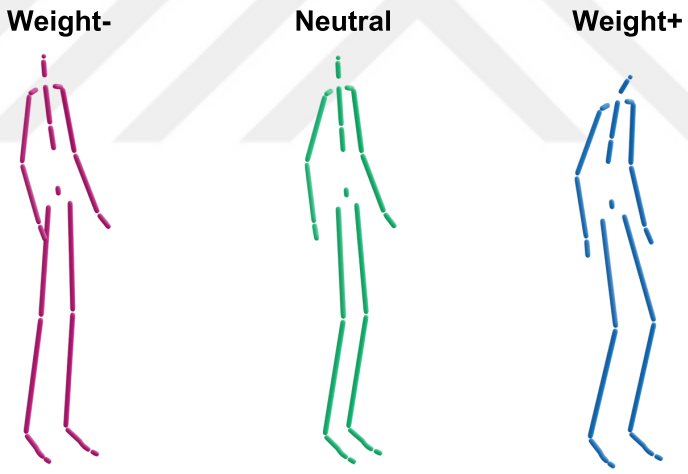

Figure 3.5: The impact of the Weight factor.

3.3.3 *Time*

The *Time* effort parameter represents the variation of movement over time. Motion is categorized into two types: *quick* and *sustained*. In our study, we applied the method described in [6] to influence movement within the time domain.

In the original study, the effect of *Time* parameter on motion was achieved by accelerating or decelerating the animation. The playback speed of each frame was irregularly modified to preserve the naturalness of movement. First, the average

speed of the hand bones in each frame was computed, and the frames were then ranked from relatively fast to slow. The playback speeds of these ordered frames were determined based on the *Time* factor, f_T. When $f_T = 1$, the playback speed ranged from $[1, 2]$, and when $f_T = -1$, the playback speed ranged from $[0.5, 1]$.

In our approach, we extend this method by incorporating both the hands and feet. For each frame difference, ΔT, we calculate the total average displacement by averaging the displacements of the hand and foot bones during ΔT. Frames are then reordered based on average displacement, and keyframes are shifted accordingly. We use f_T to compute the speed factor, f_s, to determine the new keyframe positions (cf. Equation 3.8). Figure 3.6 shows the effect of the *Time* factor in positive and negative directions.

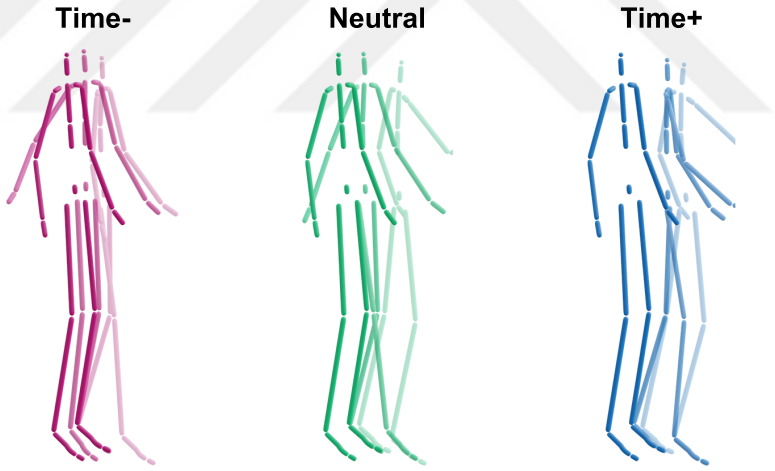

Figure 3.6: The impact of the *Time* factor.

$$f_s = \begin{cases} f_T + 1, & \text{if } f_T \geq 0, \\ \dfrac{1}{|f_T| + 1}, & \text{otherwise} \end{cases} \tag{3.8}$$

The keyframe shift amount, s_k, is the inverse of f_s since a higher speed factor results in a smaller shift. For each displacement, we calculate incremental changes. The minimum increment, inc_{min}, is $\frac{|1-k_s|}{N}$, where N is the total number of displacements. The new shift for each keyframe, $S(k_i)$, is calculated as:

$$S(k_i) = s_k + i \cdot inc_{min}. \tag{3.9}$$

3.3.4 *Flow*

The last of the effort parameters, *Flow*, represents the continuity of movement within the process. A movement is categorized as either *bound* or *free* under the *Flow* parameter. We apply two different methods to modify the movement in the direction of *Flow*.

The movement is restricted when the *Flow* factor f_F is negative. We applied the *Decimate Keyframes* operation to restrict the movement, a built-in feature in Blender. This process reduces the number of keyframes while maintaining the overall animation curve as much as possible. It helps simplify the animation data without significantly altering the content of the movement. The function takes a *Decimation Factor* f_{dec} as input. This factor can be any real number between $[0, 1]$. As the Decimation Factor approaches 1, the animation becomes more simplified. Based on our trials with several different animations, we concluded that the $[0, 0.9]$ range works best for this factor. So, f_{dec} can be calculated as $0.9 \cdot f_F$.

When f_F is positive, the animation is allowed to move more freely. In this context, as in the studies of Durupınar et al. [5] and Sonlu et al. [6], a certain amount of rotation is added to each bone in the skeleton at random keyframes to demonstrate the uncontrollability of the movement.

A function generates keyframe positions based on f_F to select these random keyframes. The function calculates a range for the intervals between keyframes: if f_F is close to 1, the range is narrower, making keyframes more closely spaced. Specifically, the range is determined by a minimum of 5 frames and a maximum of 30 frames, with the exact interval varying depending on f_F. The function then randomly selects increments within this computed range to determine the positions of the keyframes throughout the animation. The ratio of the selected

random keyframes and the amount of rotation added are directly proportional to the determined f_F.

3.4 Data-driven Approach

Our data-driven approach consists of two parts: *data processing* and *the deep learning architecture.*

3.4.1 Data Processing

We labeled the Bandai-Namco dataset according to the OCEAN model. After the annotations and the filtering stages, we were left with 128 labeled animations. The rest of the data was not annotated; we had 3,077 animations. However, this number of examples was insufficient for effective training. In our model, the reconstruction of personality traits relies on having a diverse dataset to ensure that the model can generalize well across different personality dimensions. To achieve tangible results, we needed a more substantial amount of labeled data.

To increase the total amount of augmented data, we first divided each animation into segments of 128 frames. If an animation was shorter than 128 frames, we repeated the last frame to pad the sequence until it reached 128. Additionally, we applied a sliding window approach to generate overlapping animations, further increasing the data. The window progression was controlled by the "overlap ratio." In our case, we used an overlap ratio of 0.75, meaning each new segment includes the last 96 frames from the previous segment, with 32 new frames added. We also introduced the "drop ratio," determining when the last segment should be discarded. We set the drop ratio to 0.67, meaning it would be dropped if the new segment contained fewer than 86 non-padded frames.

After creating the overlapping animations, the total number of animations increased to 7,171, with 390 annotated. We then split the dataset into training

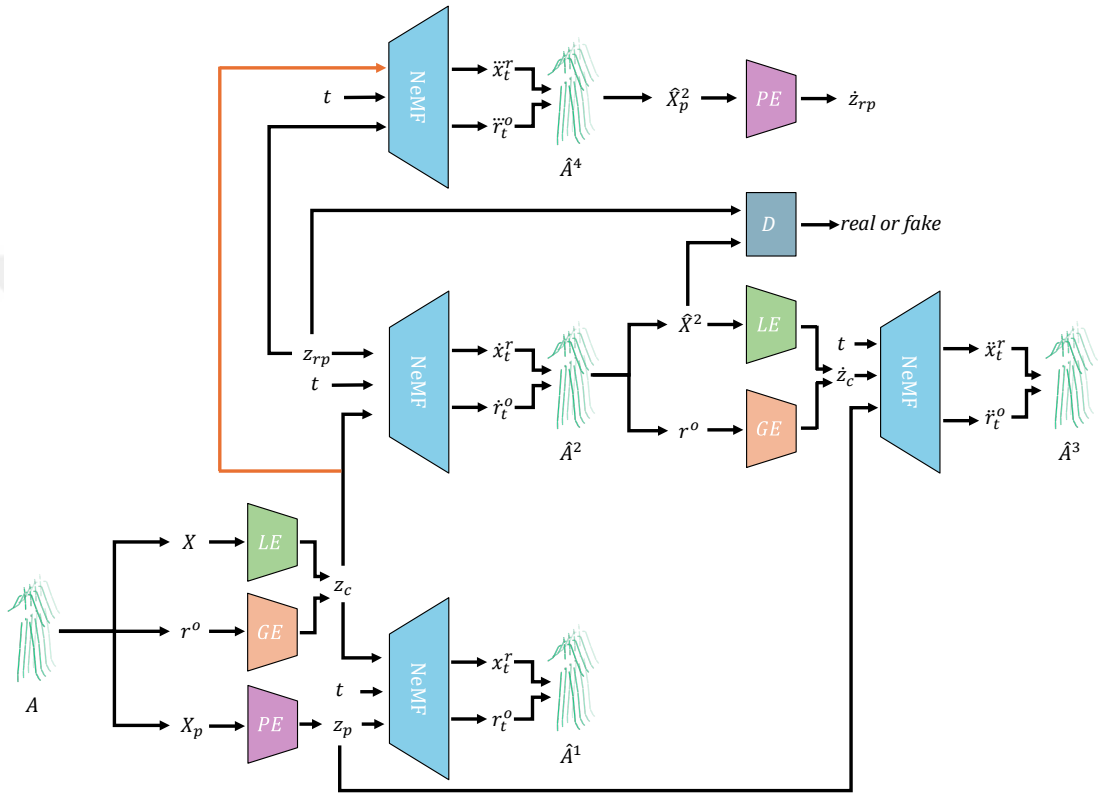

Figure 3.7: Overview of our personality transfer model, P-GeNeMF. The network takes Animation A and aims to reconstruct Animation $\hat{A}^1$ by learning the content latent variable (z_c) and personality latent variable (z_p). The local input features (X) are passed through the Local Encoder (LE), and global features (r^o) are passed through the Global Encoder (GE), resulting in the local latent variable (z_l) and the global latent variable (z_g), respectively. These two latent variables are combined to construct z_c. The personality input features X_p are passed through the Personality Encoder (PE), resulting in the personality latent variable (z_p). A random personality variable (z_{rp}) is passed through another $NeMF$ module with z_c to construct Animation $\hat{A}^2$, which essentially contains the content of $\hat{A}^1$ with a different personality. The discriminator D ensures $\hat{A}^2$ is real. A third $NeMF$ generates $\hat{A}^3$ by getting the content variable $\dot{z}_c$ from $\hat{A}^2$ and personality variable z_p to ensure consistency between $\hat{A}^1$ and $\hat{A}^3$. Finally, a fourth $NeMF$ (placed at the top of the figure) is used to ensure consistency between z_{rp} and $\dot{z}_{rp}$. The variable z_c is passed as detached, which cuts the gradient flow to LE and GE. This ensures that only PE is trained to learn $\dot{z}_{rp}$ without affecting z_c.

and validation sets using a 90/10 ratio. During this process, we ensured that no segments from the same animation were present in both sets simultaneously to avoid inconsistencies and data leakage during training.

3.4.2 Deep Learning Architecture

We employed the NeMF model, as outlined by He et al. [27], as the backbone of our architecture. The architecture is extended to form a semi-supervised personality-driven generative model called P-GeNeMF, representing motion as content (z_c) and personality (z_p). This approach incorporates personality traits (OCEAN model) into the latent variable to condition the generation of animations. Figure 3.7 depicts the overview of our architecture.

3.4.2.1 Personality Encoder

We introduced a new personality encoder based on the existing local and global encoders of generative NeMF to extract the personality latent variable (z_p). The input X_p to the personality encoder includes LMA-based features. We calculate the personality features as follows:

- *Space Parameter:* Calculated using L_2 distances (D_t^p) between key joints, including
  - hands,
  - lower arms,
  - upper arms,
  - feet,
  - lower legs,
  - upper legs,
  - head and each hand,
  - hips and head,
  - hips and each hand, and

  - each lower arm and each hand.

- *Weight Parameter:* Derived from angles (A_t^r) between joints, such as
  - l/r hand – l/r lower arm – l/r upper arm,
  - l/r foot – l/r lower leg – l/r upper leg,
  - l/r lower arm – l/r upper arm – l/r shoulder,
  - head – neck – spine,
  - chest – spine – hips,
  - left hand – hips – right hand,
  - left upper arm – hips – right upper arms, and
  - left foot – hips – right foot.

- *Time and Flow parameters:* We calculated joint velocities ($\dot{x}_t^p$) and angular accelerations ($\ddot{x}_t^r$), as in [7] for these respective parameters.

We refer to the joint distances and angles as joint-specific personality features, while the velocities and angular accelerations of joints are called local personality features. The joint-specific personality features are processed through four residual blocks, as used in the global encoder of the generative NeMF model. Each of these residual blocks consists of 1D convolutional layers. Similarly, the local personality features are passed through four skeleton residual layers, also following the local encoder of the generative NeMF architecture. Each output feature vector has 256 dimensions. These vectors are concatenated to create a 512-dimensional vector, which is then passed through a linear layer to generate the 5-dimensional personality latent variable z_p.

3.4.2.2 Decoding and Reconstruction Losses

After generating the personality latent variable, we pass t, z_c, and z_p through a NeMF decoder, and the outputs joint rotations (x_t^r) and (r_t^o) are generated. Similarly, as in generative NeMF, these outputs are used in the reconstruction loss ($\mathcal{L}_{rec}$) for the animation.

$$f(t, z_c, z_p) \rightarrow (x_t^r, r_t^o) \rightarrow \hat{A}^1, \tag{3.10}$$

where x_t^r and r_t^o indicate the reconstructed motion's joint rotations and root orientation, respectively. Here, z_c comprises a local latent variable (z_l) and a global latent variable (z_g). These are generated by the local and global encoders, respectively, as described in Section 2.6.

During training, leveraging animations with labeled personality traits, a personality reconstruction loss $\mathcal{L}_{pers}$ was used. These labels served as supervision to optimize the model's ability to learn and replicate the target personality features from the animation data. As some animations were not labeled, we applied masking to hide the unlabeled ones while calculating the $\mathcal{L}_{pers}$.

3.4.2.3 Semi-supervised Learning

From our experiments, we found that using only a small amount of supervised data did not provide consistent results. To improve this, we adopted a semi-supervised approach. We jointly trained the model using both labeled and unlabeled data. The personality encoder was trained to accurately predict the ground truth personality traits for the labeled data. For the unlabeled data, the personality encoder learned the distribution of the personality traits.

We introduced the generation of a secondary animation $\hat{A}^2$ to allow our model to explore more animations and personality traits. This new animation combines the content latent variable z_c extracted from A with a randomly sampled personality latent variable z_{rp}, where the OCEAN traits are uniformly sampled from the range $[-3, 3]$.

$$f(t, z_c, z_{rp}) \rightarrow (\dot{x}_t^r, \dot{r}_t^o) \rightarrow \hat{A}^2. \tag{3.11}$$

3.4.2.4 Adversarial Training

To ensure the realism of the generated animation $\hat{A}^2$ with the fake personality, we applied an adversarial strategy. We have used the discriminator D to decide whether $\hat{A}^2$ should be classified as real or fake. We trained D using the ground truth animation A and the reconstructed animation $\hat{A}^2$. Here, the ground truth animation A is the *real* sample while $\hat{A}^2$ was the *fake* one. Since some of our data lacked labels, instead of training the discriminator with ground truth personality labels, we utilized the learned personality representations, z_p. We used the Wasserstein-GAN with gradient penalty algorithm [46] while training D.

The structure of D is similar to the local encoder in NeMF. It takes local motion features ($\hat{X}^2$) that are reconstructed from $\hat{A}^2$ and z_{rp} as inputs. The local motion features are processed directly through four residual blocks constructed from skeleton convolution and skeleton pooling layers with activation function PReLU. The output of these layers is then concatenated with z_{rp}, and the concatenated result is passed through several linear layers, each followed by activation layers. The final output indicates whether $\hat{A}^2$ is real or fake.

3.4.2.5 Consistency and Further Synthesis

In addition to leveraging the discriminator, we passed the generated output $\hat{A}^2$ through the NeMF module and synthesized a new animation, denoted as $\hat{A}^3$. This new animation was generated by combining the content latent variable $\dot{z}_c$ of $\hat{A}^2$ with the original personality latent variable z_p. This process aimed to enforce consistency between the input animation and $\hat{A}^3$, ensuring that the original animation's key attributes (content and personality) are preserved through the transformation.

$$f(t, \dot{z}_c, z_p) \rightarrow (\ddot{x}_t^r, \ddot{r}_t^o) \rightarrow \hat{A}^3. \tag{3.12}$$

We used an additional NeMF module to ensure the consistency of z_{rp}. We input a detached z_c and z_{rp} into this module to generate $\hat{A}^4$. Next, $\hat{A}^4$ was passed through the personality encoder, producing a new personality variable, $\dot{z}_{rp}$.

3.4.2.6 Loss Functions and Total Loss

We summarize the loss functions in the sequel. The total loss $\mathcal{L}$ is used to train the P-GeNeMF model.

Motion Reconstruction Loss: The motion reconstruction loss ($\mathcal{L}_{rec}$) are the weighted sums of $\mathcal{L}_{pos}$, $\mathcal{L}_{rot}$ and $\mathcal{L}_{ori}$ as described in Section 2.6.1. We calculate these losses by calculating the input animation A and the generated animation $\hat{A}^1$. λ_{pos}, λ_{rot}, λ_{ori} are selected as 20, 2, and 2, respectively.

$$\mathcal{L}_{rec} = \lambda_{pos}\mathcal{L}_{pos} + \lambda_{rot}\mathcal{L}_{rot} + \lambda_{ori}\mathcal{L}_{ori}. \tag{3.13}$$

KL Divergence Loss: We only used KL Divergence Loss $\mathcal{L}_{KL}$ to ensure the content variable z_c follows the standard normal.

$$\begin{aligned} \log p(x_r, r_o) \geq\ & \mathbb{E}_{q_{\theta_1}, q_{\theta_2}} \left[\log p(x_r, r_o \mid z_l, z_g)\right] \\ & - D_{KL}(q_{\theta_1}(z_l \mid X) \parallel p(z_l)) \\ & - D_{KL}(q_{\theta_2}(z_g \mid r_o) \parallel p(z_g)). \end{aligned} \tag{3.14}$$

Personality Reconstruction Loss: Because not all animations were labeled, we used a mask M to mask out the unlabeled animations to find the personality reconstruction loss $\mathcal{L}_{pers}$:

$$\mathcal{L}_{\text{OCEAN}} = \max\left(\frac{1}{N}\sum_{i=1}^{N}(P^{(i)} - z_p^{(i)})^2, m\right) - m, \tag{3.15}$$

where P is the personality label, z_p is the personality latent variable, m is a margin, and N is the number of samples inside the batch. The overall loss $\mathcal{L}_{pers}$ is then averaged over the batch, considering the mask:

$$\mathcal{L}_{pers} = \frac{\sum_{i=1}^{B} \left(\mathcal{L}_{\text{OCEAN}}^{(i)} \cdot M^{(i)} \right)}{\max \left(1, \sum_{i=1}^{B} M^{(i)} \right)}. \tag{3.16}$$

Motion Consistency Loss: The motion consistency loss ($\mathcal{L}_{cons}$) is similar to motion reconstruction loss ($\mathcal{L}_{rec}$). Instead of using the generated animation $\hat{A}^1$, we have used $\hat{A}^3$ to compare it with the input animation A. λ_{cpos}, λ_{crot}, λ_{cori} are selected as 10, 1, and 1, respectively.

$$\mathcal{L}_{cons} = \lambda_{cpos}\mathcal{L}_{cpos} + \lambda_{crot}\mathcal{L}_{crot} + \lambda_{cori}\mathcal{L}_{cori}. \tag{3.17}$$

Personality Consistency Loss: The consistency loss ($\mathcal{L}_{pcons}$) between the random personality latent z_{rp} and the constructed personality latent $\dot{z}_{rp}$ from the generated animation $\hat{A}^4$. Similar to $\mathcal{L}_{pers}$, we have used Mean Squared Error (MSE). N stands for the total number of samples in the dataset:

$$\mathcal{L}_{pcons} = \max \left(\frac{1}{N} \sum_{i=1}^{N} (z_{rp}^{(i)} - \dot{z}_{rp}^{(i)})^2, m \right) - m. \tag{3.18}$$

Content Consistency Loss: We also experimented with content consistency loss ($\mathcal{L}_{ctcons}$) between the content latent z_c and the constructed content latent $\dot{z}_c$ from the generated animation $\hat{A}^2$. For both local and global variables in the content variables, we used the MSE loss, and again, N stands for the total number of samples in the dataset. Both loss weights, λ_{cl} and λ_{cg}, are selected as 1:

$$\mathcal{L}_{cl} = \frac{1}{N}\sum_{i=1}^{N}(z_l^{(i)} - \dot{z}_l^{(i)})^2 \quad \mathcal{L}_{cg} = \frac{1}{N}\sum_{i=1}^{N}(z_g^{(i)} - \dot{z}_g^{(i)})^2 \tag{3.19}$$

$$\mathcal{L}_{ctcons} = \lambda_{cl}\mathcal{L}_{cl} + \lambda_{cg}\mathcal{L}_{cg}. \tag{3.20}$$

Generator Loss: This loss is simply the negative mean of the output of discriminator D because we use the Wasserstein-GAN with gradient penalty algorithm (W-GAN GP) [46] while training D. For the total number of N samples in the dataset, the generator loss is given by

$$\mathcal{L}_{gen} = -\frac{1}{N}\sum_{i=1}^{N} D(\hat{X}^2, z_{rp})^{(i)}. \tag{3.21}$$

Total Loss: The total loss is the weighted sum of all loss functions. The weights that are present inside the losses are not displayed again. λ_{pers}, λ_{pcons}, λ_{gen}, and λ_{KL} are selected as 0.3, 3, 1, and 10^{-5}, respectively:

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda_{pers}\mathcal{L}_{pers} + \mathcal{L}_{cons} + \lambda_{pcons}\mathcal{L}_{pcons} + \mathcal{L}_{ctcons} + \lambda_{gen}\mathcal{L}_{gen} + \lambda_{KL}\mathcal{L}_{KL} \tag{3.22}$$

Discriminator Loss: We trained D using the input animation A and the reconstructed animation $\hat{A}^2$. For each loop of training P-GeNeMF, we trained D twice. The discriminator loss $\mathcal{L}_{disc}$ is calculated as follows:

$$\mathcal{L}_{\text{real}} = \mathbb{E}[D(A)], \quad \mathcal{L}_{\text{fake}} = \mathbb{E}[D(\hat{A}^2)], \tag{3.23}$$

$$\mathcal{L}_{\text{GP}} = \lambda_{\text{GP}} \cdot \mathbb{E}\left[(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2\right], \tag{3.24}$$

$$\mathcal{L}_{disc} = \mathcal{L}_{\text{fake}} - \mathcal{L}_{\text{real}} + \mathcal{L}_{\text{GP}}. \tag{3.25}$$

where $\hat{\mathbf{x}}$ is a random interpolation between A and $\hat{A}^1$. λ_{GP} is the loss weight and is set as 10.

3.4.2.7 Training Details

We train our model on an HP® Proliant DL380 Gen9 server with 2 Intel® Xeon® E5-2620 v4 CPUs @ 2.10GHz, 64GB memory and a 16GB NVIDIA Tesla P100 GPU. We have used Python 3.9.13 and PyTorch 1.12.1 with CUDA 11.6. Our model is built upon the code from the NeMF: Neural Motion Fields repository of He et al., available at `https://github.com/c-he/NeMF/tree/main`. We employed the Adam optimizer to train both P-GeNeMF and the discriminator, as described in [27]. The learning rate for both optimizers was set to 0.0001. As discussed in the discriminator loss section, p-GeNeMF was trained for 300 epochs, while the discriminator was trained for twice as many iterations. P-GeNeMF optimizer had a weight decay of 0.0001, while the discriminator optimizer did not have any weight decay.

Chapter 4

Experimental Evaluation

4.1 User Studies

We conducted three user studies to evaluate both of our methods. The first and third studies focused on transferring personality traits from Animation B to A. In the second study, we focused on changing the personality traits of each motion, either in a positive or negative direction. Each study design and the results of them are explained in the following subsections. The screenshots for each study are provided in Appendix A.

4.1.1 Study 1

In Study 1, we aim to demonstrate the effectiveness of our handcrafted modulation tool and P-GeNeMF in influencing the personality of Animation A when transferring personality traits from Animation B. We selected ten animation pairs. Six of these pairs shared the same content, where both animations either performed a dynamic action or remained in a stable motion. For the other four pairs, we selected animations with differing content motions. We transferred the personality traits from B to A for all ten pairs and then reversed the transfer

direction, resulting in 20 animations.

For each task, the participant rated the perceived personality of the samples using the sliders below each animation. Sliders measure the OCEAN personality traits in five dimensions. Additionally, the realism (human-likeness) of each motion is also rated. The study has twenty tasks, including four randomly arranged animations for each task: Animation A (content motion), the result of the handcrafted modulation tool, P-GeNeMF, and Animation B (style motion). Additionally, for each participant, the tasks are shown in random order.

The participants could also rotate the camera along the vertical axis and zoom in or out. This feature was crucial for participants to explore the animations in more depth, as some differences or details might not be immediately noticeable when viewing the animation from a static perspective.

The P-GeNeMF model could transfer personalities between motions by switching their personality latent codes (z_p). We followed a simple approach to modulate the motion using the handcrafted tool. Since the personality labels were not explicitly defined, we manually adjusted each LMA parameter to match Animation A's effort parameters with those of Animation B.

4.1.2 Study 2

In Study 2, we attempted to modify each personality trait in both positive and negative directions for Animation A to evaluate the effectiveness of the models. For each personality trait T, we selected two animations with neutral values for the trait. These two animations were then altered in the $T+$ and $T-$ directions, resulting in 20 animations for both models.

Similar to Study 1, participants rated the perceived personality of the samples using sliders measuring the OCEAN personality traits across five dimensions and the realism (human-likeness) of each motion. Twenty tasks were presented in this study, each containing three randomly arranged animations: Animation A

(GT), the result of the handcrafted modulation tool, and the result of P-GeNeMF. Additionally, for each participant, the tasks are shown in random order. Again, as in the previous study, the cameras were rotatable, allowing participants to explore the animations from different angles.

The results of the P-GeNeMF model were generated by preserving the content latent code (z_c) and modifying the target personality trait by either -3 or 3, depending on the direction. We employed the Laban Effort - OCEAN correlation as described in [5] to modulate the motion using the handcrafted tool. The correlations can also be seen in Table 4.1

Table 4.1: The Laban Effort - OCEAN correlations based on the findings of [5]. A "+" indicates a positive correlation, while a "−" represents a negative correlation between the Effort Parameter and the corresponding OCEAN trait. Empty cells show that the effort parameter does not affect the trait.

	Space	Weight	Time	Flow
Openness	−			−
Conscientiousness	+		−	+
Extraversion	−		+	−
Agreeableness		−	−	
Neuroticism	−		+	−

4.1.3 Study 3

The final study focused on the content, style, and personality resemblance of character animations. Each participant was asked to compare Animation A (content motion), the result of P-GeNeMF, and Animation B (personality motion). They evaluated the CENTER animation (P-GeNeMF) by comparing it with A and B. Each ground truth motion was randomly arranged.

For the animation samples, we used the same ones from Study 1. There were twenty tasks, and for each task, the participants adjusted horizontal sliders to indicate their choices for each criterion: content, style, and personality. Like in

studies 1 and 2, the tasks are shown in random order for each participant.

4.2 Results and Discussion

4.2.1 Study 1

A total of 44 unique users participated in Study 1. Between 22 and 23 different participants rated each sample. Initially, we did not have personality labels for the pairs of animations (Animation A and B). Therefore, we used the OCEAN personality values obtained from the experiment for each pair. After collecting these values, we categorized the pairs into groups based on individual personality traits. The OCEAN traits of the animation pairs determined each group. Animations were classified as low (L), neutral (N), or high (H) for each trait. The groups were created from all possible combinations of these classifications (e.g., L-L, L-N, and L-H); there can be at most nine groups for each OCEAN trait.

Table 4.2: Sample classifications based on OCEAN traits. Each sample is categorized as 'Low,' 'Neutral,' or 'High' for a specific trait, depending on where its value falls within the defined range. The values for each trait are assigned to one of these three categories based on the range boundaries, where r_s and r_e represent the start and end of the range, respectively.

Traits	O		C		E		A		N	
Classes	r_s	r_e	r_s	r_e	r_s	r_e	r_s	r_e	r_s	r_e
Low	-1.405	-0.260	-1.290	-0.281	-2.040	-0.507	-0.750	0.265	-1.603	-0.777
Neutral	-0.260	0.885	-0.281	0.727	-0.507	1.031	0.265	1.280	-0.777	0.048
High	0.885	2.031	0.727	1.735	1.031	2.570	1.280	2.296	0.048	0.874

Table 4.2 provides the details of the classification ranges for each OCEAN trait. These classifications result in varying numbers of groups for each trait. There are six distinct groups for openness and five groups for extraversion. Conscientiousness, agreeableness, and neuroticism each have seven distinct groups.

To determine the ranges for each OCEAN trait, we first calculated the mean values of Animation A and B for each sample. From these mean pairs, we identified the minimum and maximum values. These values were then divided into three equal parts to define the range for each sample classification.

From the same table, we can also observe that each OCEAN trait has distinct classification ranges. These differences can be highlighted by calculating the gap between the r_e value of 'High' and the r_s value of 'Low.' Among the traits, extraversion (E) has the largest range, while neuroticism (N) has the smallest. This variability shows that analyzing the samples by these groups was necessary to accurately reflect the variations in each trait and provide more reliable results.

Each pairwise comparison figure presents distinct groups for the specific personality trait. In these figures, each group displays the distribution of generated results, including content motion (A) and personality motion (B). For each generated result, we aim to observe the mean value between A and B, with a preference for values closer to B. Between the two generated results, the one closer to B for a given personality trait can be considered more successful in that context.

The analysis of the openness trait, as shown in Figure 4.1, reveals distinct performance differences between the data-driven model (NeMF) and the handcrafted tool (HC). In the "High-Low," "High-Neutral," and "Neutral-Neutral" groups, NeMF successfully transfers the target personality (B), outperforming HC. The handcrafted tool is notably effective only in the "High-Low" group. The pairwise comparisons (cf. Table 4.3) support these findings, where the significance value (p) confirms that the mean differences between B-NeMF and B-HC in the "High-Low" group are not coincidental. This observation suggests that both models perform well in this group, though the data-driven approach is generally more effective.

For the conscientiousness trait, displayed in Figure 4.2, both models exhibit reasonable performance, particularly when transferring a higher trait level to a lower one. In the "Low-Neutral" group, both data-driven and handcrafted methods successfully raise the trait level, indicating that a smaller gap between

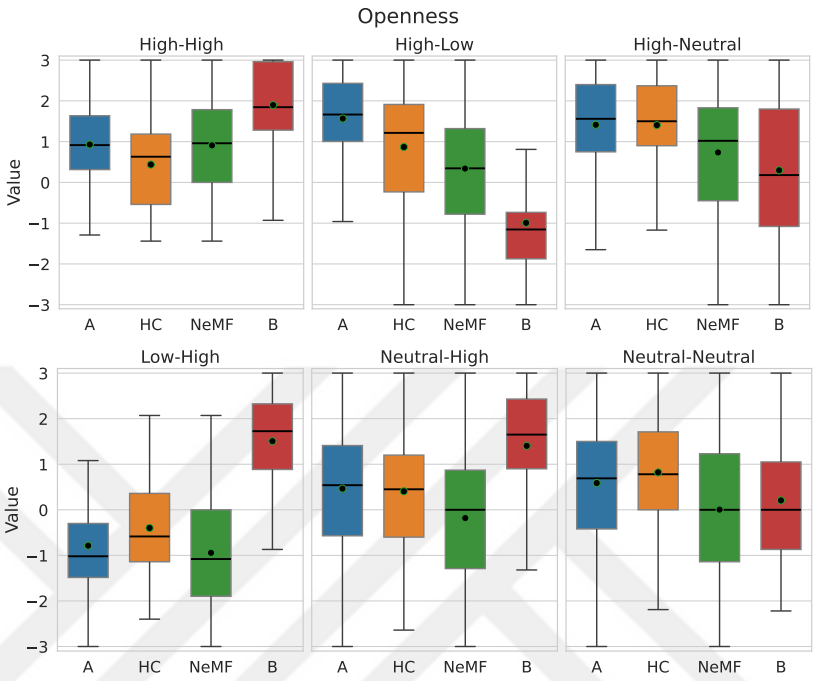

Figure 4.1: Box-plots for pairwise comparisons of the openness groups (Study 1).

extremes enhances transferability. The pairwise comparisons demonstrate significant mean differences between the target personality distribution and the handcrafted method in the "Neutral-Low," and "High-Low," groups, showing that both models perform effectively in these cases.

When examining the extraversion trait in Figure 4.3, it is clear that both data-driven and handcrafted methods perform well in transferring the trait to a lower level across most groups. This observation is validated by the p values in the pairwise comparisons, which suggest significant differences in all groups except the "Neutral-Neutral" case for the data-driven approach. Interestingly, in the "Neutral-Neutral" group, the data-driven approach surpasses the target distribution, suggesting that while the model performs well, adjustments may be necessary in cases where the target trait remains constant. This problem could be addressed by fine-tuning the scales between the two personality latent vectors to minimize such discrepancies.

Figure 4.4 shows that the results for the agreeableness trait follow a similar trend to those seen with other personality traits, where both models perform well

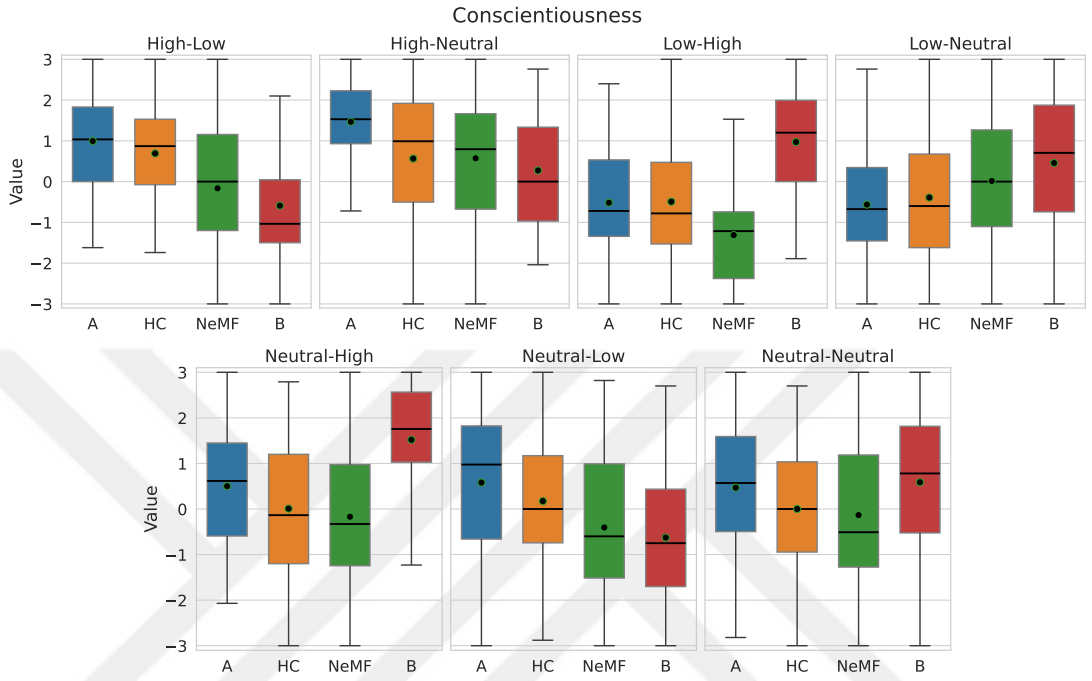

Figure 4.2: Box-plots for pairwise comparisons of the conscientiousness groups (Study 1).

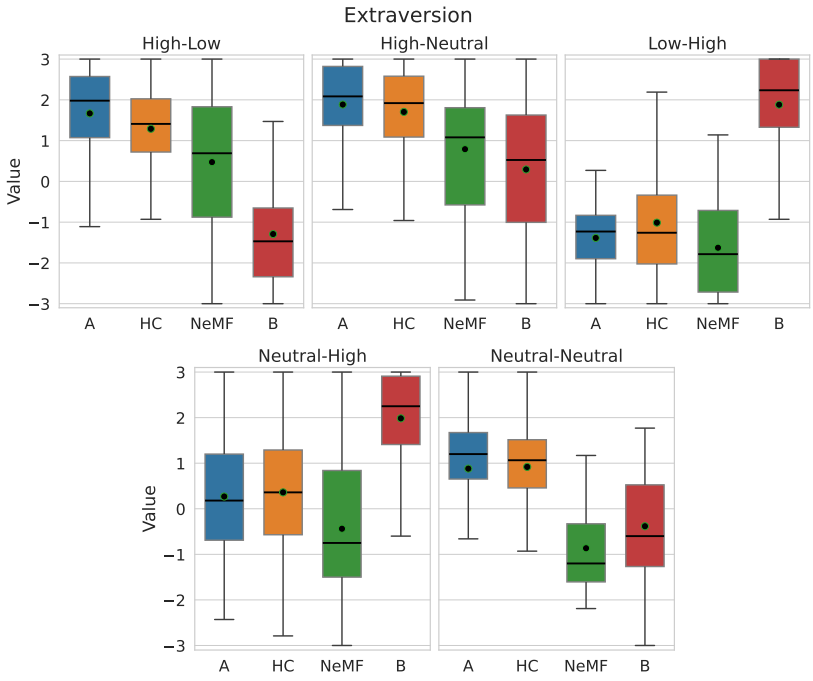

Figure 4.3: Box-plots for pairwise comparisons of the extraversion groups (Study 1).

when tasked with decreasing the trait. The data-driven method appears more effective than the handcrafted tool in these cases. The p values from the pairwise comparisons indicate that the differences between the target personality and the data-driven approach are statistically significant, except in the "High-Neutral" group. This observation suggests that the data-driven consistently produces more accurate personality transfers, emphasizing its potential over the handcrafted method in most scenarios.

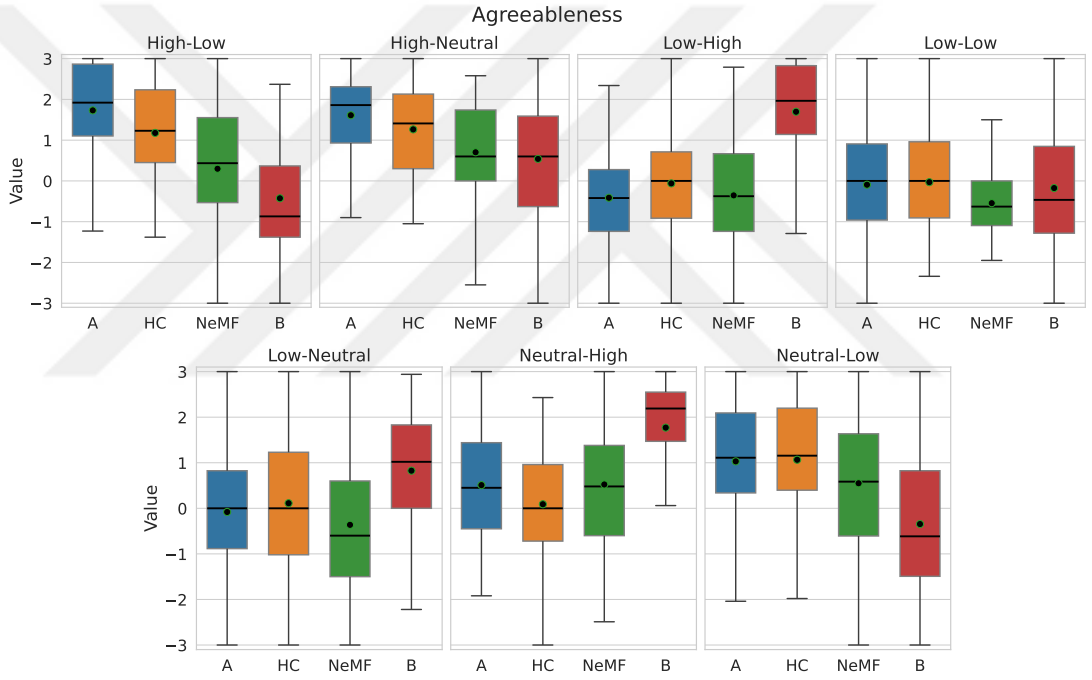

Figure 4.4: Box-plots for pairwise comparisons of the agreeableness groups (Study 1).

Figure 4.5 shows that increasing the neuroticism trait can be done more accurately than decreasing it for both methods. The pairwise comparisons confirm this, as the p values show significant mean differences between the generated results and the target personality in the "Low-High" group. Additionally, the figure shows that in the "Low-Neutral" group, the handcrafted method more closely aligns with the target personality. At the same time, the data-driven model overshoots the intended mean, again highlighting a potential need for adjustment in the data-driven approach when transferring traits to more moderate levels.

Based on the findings across all traits, a consistent pattern emerges regarding

Table 4.3: Pairwise comparisons of OCEAN trait groups and the corresponding ANOVA statistics for Study 1. Bold values show $p < 0.05$. The mean differences are calculated by subtracting the second pair (Y) from the first (X), where the paired label is shown as X-Y.

Openness

Pairs	A-B		A-HC		A-NeMF		B-HC		B-NeMF		HC-NeMF	
Groups	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p
High-Neutral	-1.117	**<.001**	-0.009	.999	-0.678	**<.001**	1.108	**<.001**	0.438	.057	-0.670	**<.001**
High-High	0.967	**.040**	-0.492	.516	-0.022	**<.001**	-1.459	**<.001**	-0.989	**.034**	0.470	.554
High-Low	-2.558	**<.001**	-0.698	**.018**	-1.226	**<.001**	1.860	**<.001**	1.331	**<.001**	-0.529	.116
Neutral-High	0.938	**<.001**	-0.059	.984	-0.647	**<.001**	-0.997	**<.001**	-1.585	**<.001**	-0.588	**.002**
Low-High	2.293	**<.001**	0.388	.429	-0.157	.928	-1.905	**<.001**	-2.450	**<.001**	-0.545	.147
Neutral-Neutral	-0.381	.574	0.233	.861	-0.583	.206	0.615	.167	-0.201	.905	-0.816	**.033**

Conscientiousness

Pairs	A-B		A-HC		A-NeMF		B-HC		B-NeMF		HC-NeMF	
Groups	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p
Neutral-High	1.019	**<.001**	-0.491	.198	-0.669	**.037**	-1.510	**<.001**	-1.688	**<.001**	-0.178	.890
Neutral-Low	-1.206	**<.001**	-0.406	.427	-0.986	**.002**	0.800	**.016**	0.221	.842	-0.579	.136
Neutral-Neutral	0.119	.962	-0.468	.222	-0.602	.067	-0.587	.078	-0.721	**.018**	-0.134	.946
High-Neutral	-1.194	**<.001**	-0.902	**.015**	-0.895	**.016**	0.291	.761	0.298	.749	0.006	.999
High-Low	-1.584	**<.001**	-0.304	.494	-1.160	**<.001**	1.280	**<.001**	0.424	.202	-0.856	**<.001**
Low-Neutral	1.021	**.013**	0.173	.954	0.580	.300	-0.848	.055	-0.441	.545	0.407	.610
Low-High	1.488	**<.001**	0.025	**<.001**	-0.794	**.004**	-1.463	**<.001**	-2.283	**<.001**	-0.820	**.003**

Extraversion

Pairs	A-B		A-HC		A-NeMF		B-HC		B-NeMF		HC-NeMF	
Groups	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p
Neutral-High	1.716	**<.001**	0.091	.912	-0.707	**<.001**	-1.624	**<.001**	-2.422	**<.001**	-0.798	**<.001**
High-Low	-2.960	**<.001**	-0.377	.250	-1.195	**<.001**	2.582	**<.001**	1.764	**<.001**	-0.818	**<.001**
Low-High	3.268	**<.001**	0.374	.453	-0.241	.778	-2.894	**<.001**	-3.509	**<.001**	-0.615	.075
Neutral-Neutral	-1.264	**.005**	0.038	**<.001**	-1.747	**<.001**	1.302	**.004**	-0.483	.563	-1.785	**<.001**
High-Neutral	-1.592	**<.001**	-0.181	.755	-1.095	**<.001**	1.412	**<.001**	0.498	**.034**	-0.914	**<.001**

Agreeableness

Pairs	A-B		A-HC		A-NeMF		B-HC		B-NeMF		HC-NeMF	
Groups	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p
Neutral-High	1.262	**<.001**	-0.419	.382	0.015	**<.001**	-1.681	**<.001**	-1.247	**<.001**	0.434	.351
Low-High	2.110	**<.001**	0.353	.256	0.059	.990	-1.757	**<.001**	-2.051	**<.001**	-0.294	.418
High-Neutral	-1.070	**<.001**	-0.347	.565	-0.908	**.005**	0.723	**.037**	0.162	.930	-0.561	.156
Low-Low	-0.082	.992	0.063	.996	-0.452	.392	0.145	.957	-0.370	.567	-0.515	.275
Neutral-Low	-1.376	**<.001**	0.035	.999	-0.481	.214	1.411	**<.001**	0.895	**.002**	-0.516	.162
Low-Neutral	0.906	**.001**	0.192	.861	-0.282	.654	-0.714	**.019**	-1.188	**<.001**	-0.474	.212
High-Low	-2.157	**<.001**	-0.553	**.036**	-1.432	**<.001**	1.603	**<.001**	0.724	**.003**	-0.879	**<.001**

Neuroticism

Pairs	A-B		A-HC		A-NeMF		B-HC		B-NeMF		HC-NeMF	
Groups	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p	M. Diff.	Adj. p
High-Low	-1.746	**<.001**	-0.1616	.952	-0.203	.910	1.584	**<.001**	1.542	**<.001**	-0.042	.999
Low-High	1.644	**<.001**	0.421	.177	1.033	**<.001**	-1.223	**<.001**	-0.611	**.018**	0.612	**.017**
High-Neutral	-0.338	.726	0.117	.984	0.806	.066	0.454	.501	1.143	**.003**	0.689	.149
Neutral-Low	-0.646	**.024**	0.380	.337	0.823	**.001**	1.026	**<.001**	1.470	**<.001**	0.443	.207
Low-Neutral	1.032	**.016**	0.870	.059	1.382	**<.001**	-0.162	.966	0.350	.740	0.512	.449
Neutral-High	0.874	**.001**	0.060	.994	1.021	**<.001**	-0.814	**.003**	0.147	.922	0.961	**<.001**
Neutral-Neutral	-0.340	.659	-0.019	**<.001**	0.323	.695	0.321	.699	0.663	.116	0.342	.655

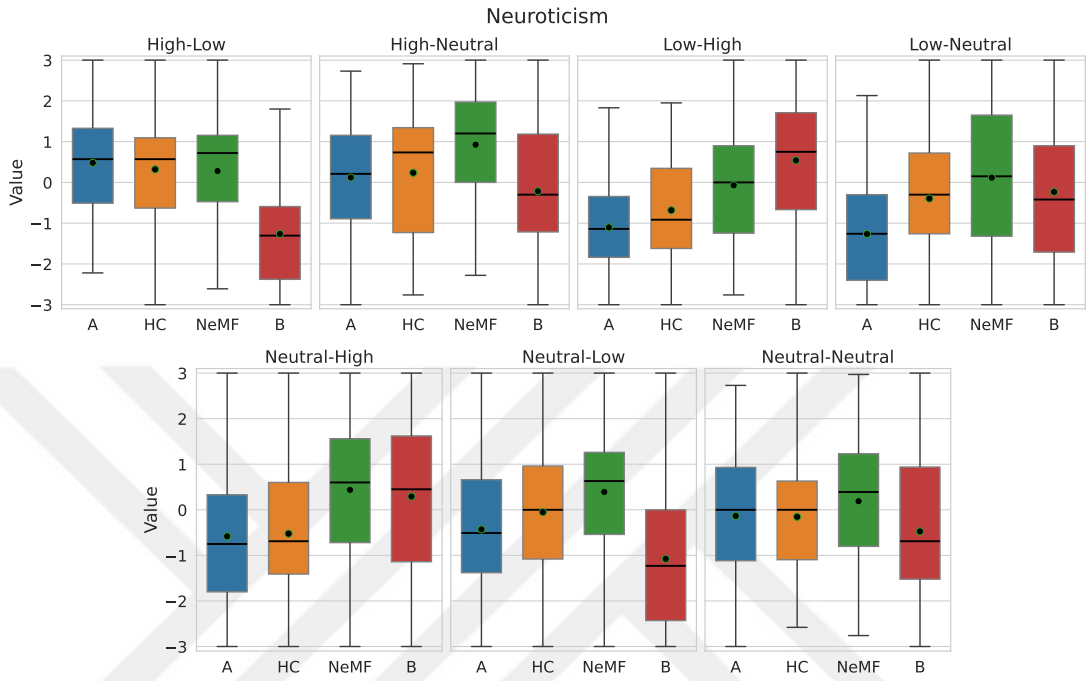

Figure 4.5: Box-plots for pairwise comparisons of the neuroticism groups (Study 1).

the effectiveness of the data-driven model and the handcrafted tool. Generally, both methods perform well when tasked with decreasing a personality trait. This observation also holds for neuroticism when viewed as emotional stability, which can be considered the inverse of neuroticism. However, the data-driven model consistently shows greater accuracy and success in most cases, particularly when significant trait changes are required. For instance, NeMF performs better when transferring traits from higher to lower levels, whereas the handcrafted tool tends to be more limited in scope.

Another recurring theme is that while the data-driven method generally produces more precise results, it sometimes overshoots the target in cases where the goal is to maintain or slightly adjust a trait, as seen in the "Neutral-Neutral" cases for extraversion. This observation suggests that while the data-driven model is more versatile and practical overall, it may require scale adjustments for moderate personality shifts to avoid exceeding the intended target.

In summary, the data-driven model outperforms the handcrafted tool across

most personality traits, particularly when significant trait shifts are needed. However, both models demonstrate strengths in reducing traits, with the data-driven being more promising in handling diverse personality transfer tasks.

Figure 4.6 demonstrates that the handcrafted model typically produces more realistic results than the data-driven approach. This discrepancy may explain why the data-driven model struggles when transferring traits from low to high levels. The handcrafted method's focus on realism may contribute to its success in maintaining a more authentic appearance, particularly in these more challenging trait transformations.

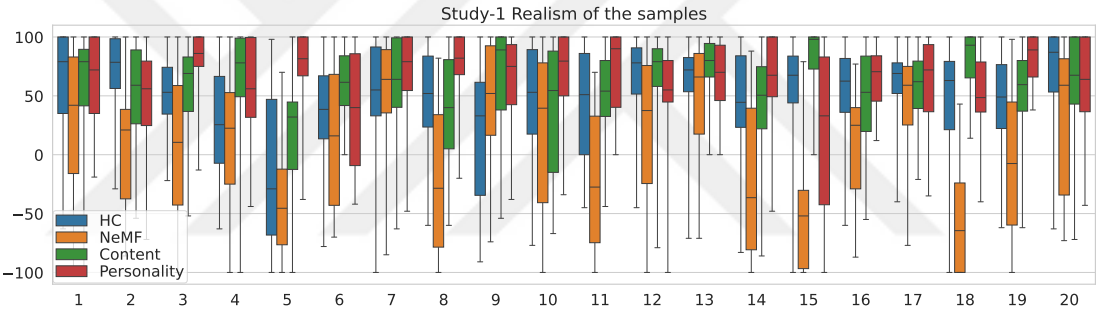

Figure 4.6: Boxplots showing the realism of generated animations in Study 1. Lower realism values indicate that users perceive the animations as less realistic.

4.2.2 Study 2

A total of 35 unique users participated in Study 2. Between 19 and 21 different participants rated each sample. Figures 4.7 to 4.11 show the trait differences for each sample between the generated animations and the ground truth. The samples are named as T_i^{sign}, where T represents the trait intended to be altered, and sign indicates the direction of the change for that trait. A positive difference value in the figures suggests that the generation method increased that trait, whereas a negative difference decreased it. A generation method can be considered successful if it modifies only the target trait without affecting the other traits.

As depicted in Figure 4.7, both models struggled to increase openness when

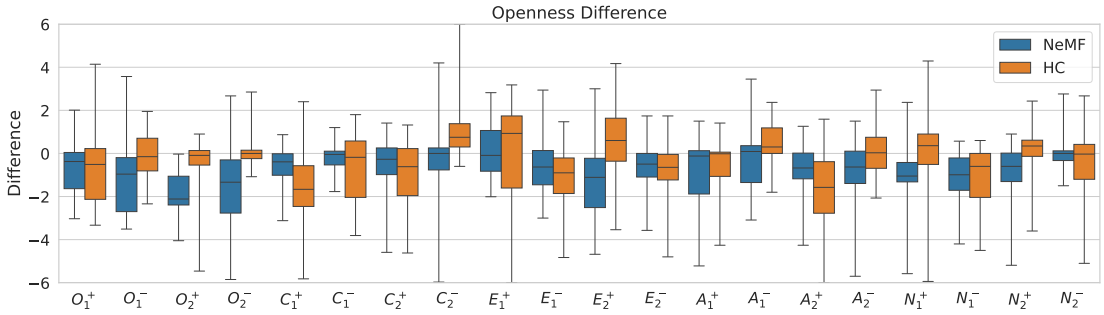

Figure 4.7: Openness differences for each sample between the generated animations and the ground truth (Study 2).

it was intended (O_i^+). However, the data-driven model outperformed the handcrafted one when the goal was to reduce openness (O_i^-). Moreover, the data-driven model usually preserved openness better than its handcrafted counterpart in cases where other traits were being modified.

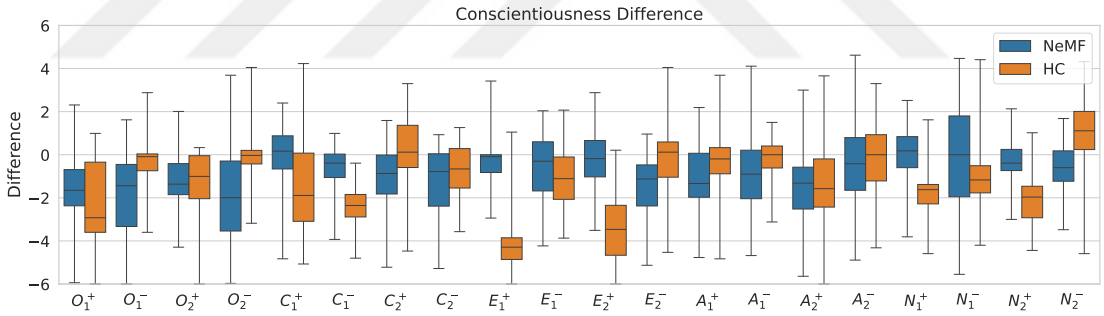

Figure 4.8: Conscientiousness differences for each sample between the generated animations and the ground truth (Study 2).

Figure 4.8 shows that the results were mixed between the two models for conscientiousness. The data-driven model outperforms at increasing conscientiousness in the first sample, while the handcrafted model performs better in the second sample. This pattern also appeared when conscientiousness was meant to be decreased. For other samples, excluding those targeting changes in extraversion and neuroticism, both the handcrafted and data-driven methods produced similar results. However, the data-driven model better preserved the conscientiousness trait for the samples where extraversion and neuroticism were modified.

Looking at Figure 4.9, handcrafted animations were more effective at changing the extraversion trait than the data-driven method in the samples aimed at

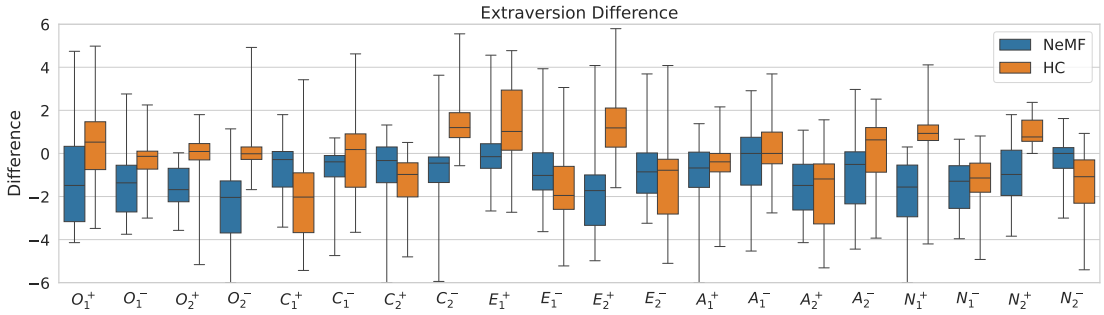

Figure 4.9: Extraversion differences for each sample between the generated animations and the ground truth (Study 2).

modifying extraversion. Likewise, for other traits intended to be modified, except for conscientiousness and neuroticism, the handcrafted model was better at preserving the extraversion trait.

The data-driven method was more effective than the handcrafted one at changing agreeableness, except for sample A_1^+ (see Figure 4.10). For other traits intended to be altered, the results for preserving agreeableness were mixed between the handcrafted and data-driven methods.

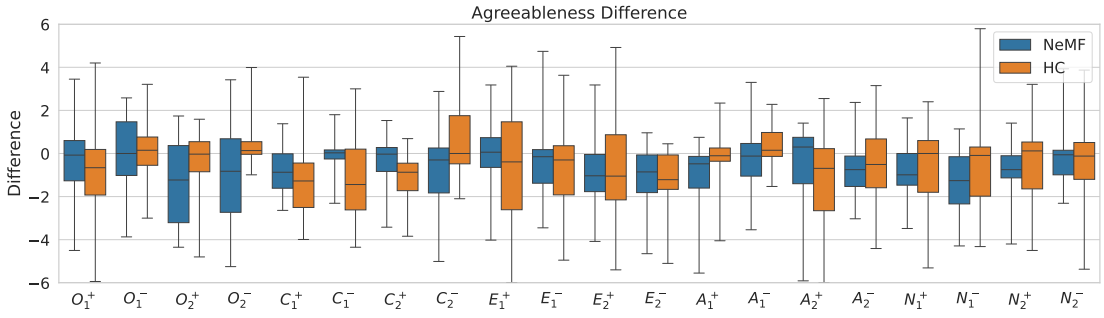

Figure 4.10: Agreeableness differences for each sample between the generated animations and the ground truth (Study 2).

Figure 4.11 shows that the handcrafted method is more effective at altering neuroticism than the data-driven method. For other traits intended to be altered, both models have mixed results when preserving the neuroticism trait. Notably, changing extraversion with the handcrafted method increases total neuroticism.

Figure 4.12 shows the boxplot depicting the realism of each animation sample generated for Study 2. Compared to the realism of generated animations in

Study 1 (cf. Figure 4.6), it is seen that only altering one personality trait reduces the result of the realism in our case.

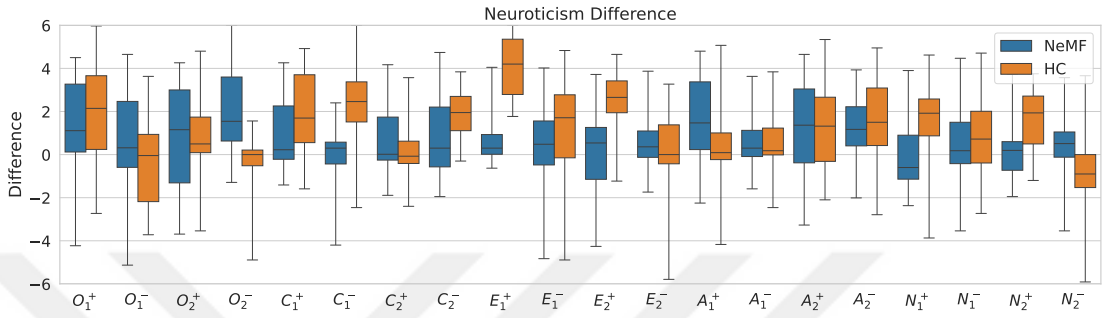

Figure 4.11: Neuroticism differences for each sample between the generated animations and the ground truth (Study 2).

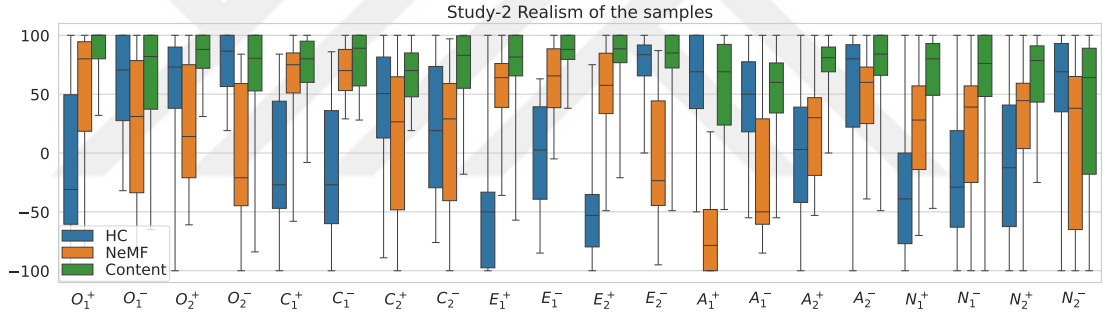

Figure 4.12: Boxplots showing the realism of generated animations in Study 2. Lower realism values indicate that users perceive the animations as less realistic.

The data-driven method was better at altering agreeableness, whereas the handcrafted method was better at changing extraversion and neuroticism. This difference can result from changing the *Space* and *Flow* factors inside the handcraft method. Referring back to Table 4.1, the effect of these factors can be more accurately represented in the handcrafted method than in the data-driven approach. Finally, in animations generated with the data-driven approach, changing the conscientiousness affected the other traits less than in the animations generated with the handcrafted method.

4.2.3 Study 3

Study 3 had a total of 27 unique participants, with each sample rated by 22 or 23 different individuals. Figure 4.13 shows that transferring personalities between animations with similar content is perceived more positively than transferring personalities between animations with different content.

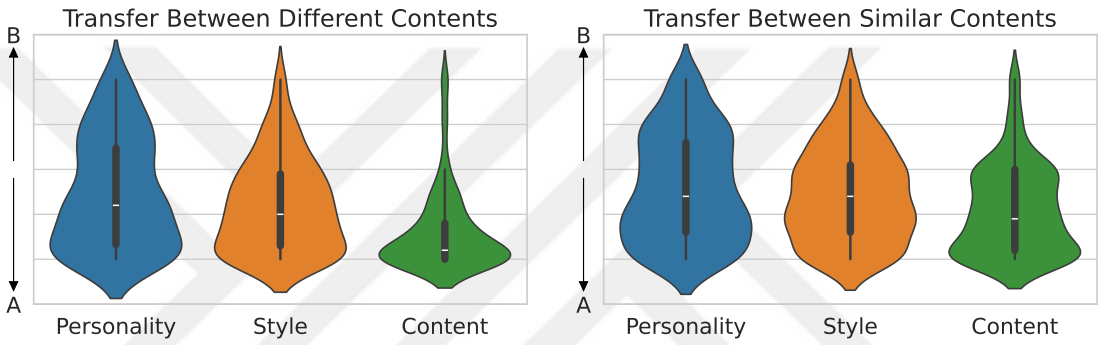

Figure 4.13: The results of Study 3. The plots show the distributions of users' choices. A distribution closer to Animation A or B indicates similarity to that animation for the given criterion. The left figure shows the distributions for personality transfer between animations with similar content, while the right shows distributions for transfers between animations with different content.

Table 4.4: Pairwise comparisons and the corresponding ANOVA statistics for Study 3. Bold values show $p < 0.05$. The mean differences are calculated by subtracting the second pair (Y) from the first (X), where the pairs are shown as X-Y.

Groups	Diff. Cont.		Sim. Cont.	
	M. Diff.	Adj. p	M. Diff.	Adj. p
Content-Personality	1.278	**<.001**	0.738	**<.001**
Content-Style	0.879	**<.001**	0.534	**<.001**
Personality-Style	-0.399	**.047**	-0.204	.310

The results depicted in Figure 4.13 and Table 4.4 and our prior user studies demonstrate that our model effectively transfers personalities without changing the style or content of the animations. While Figure 4.13 suggests that the personalities are not transferred, the findings from our previous user studies contradict this outcome. This discrepancy may be attributed to the vagueness of

the personality-related questions posed in user study three, potentially leading to varied participants' interpretations.

4.3 Ablation Study

For the ablation study, we evaluate different model versions' reconstruction, synthesis, and personality recognition capabilities. To assess motion reconstruction, we calculate the geodesic distance for joint rotations (**Rot**) and the L1 distance for joint positions (**Pos**) between the validation set and the reconstructed animation.

The synthesis capabilities of our various models are evaluated by comparing Fréchet Inception Distance (**FID**) [47] and Diversity (**Div**) [48]. The FID measures the similarity between the distributions of features extracted from generated animations and real animations. A lower FID score indicates that the generated animations are more similar to the ground truth animations. Diversity measures the variance among the generated animations, with higher scores indicating greater diversity. Like He et al. [27], we utilize a pre-trained global and local feature extractor to obtain motion features from both real and generated motions for FID and diversity measurements. This feature extractor is a regular auto-encoder with a similar architecture to GeNeMF. To compute these metrics, we first generate random motions using different P-GeNeMF models and then extract the motion features using the feature extractor. FID is calculated by measuring the Fréchet distance between the feature distributions of generated and real animations. Diversity is calculated by partitioning the features of generated animations into two sets and computing the mean Euclidean distance between these sets of feature vectors.

Additionally, we compare the performance of the personality encoder of the different models. We use the mean-squared error (MSE) of individual personality traits and their averages. We calculate the trait errors only from labeled validation animations. Lower MSE error means that the personality encoder can better

recognize the animation's personality.

We compare five different versions of our model,

- *P-GeNeMF full consistency*: Similar to our model, however, the content embedding z_c is not detached from the input of the personality consistency path and remains unchanged.
- *P-GeNeMF without personality consistency*: A version where the personality consistency loss is disabled.
- *P-GeNeMF without cycle consistency*: Our model without the cycle consistency between A and $\hat{A}^3$.
- *P-GeNeMF without discriminator*: A version of our model where the adversarial loss is removed.
- *P-GeNeMF supervised*: Similar to our model, but without semi-supervised training where only labeled data is used for training.

Table 4.5 shows the results of our ablation study. P-GeNeMF demonstrates the best motion reconstruction and synthesis results among the different variants. *P-GeNeMF without personality consistency* achieves the best performance in personality recognition but has weaker motion reconstruction and synthesis performance. *P-GeNeMF supervised* shows a good personality performance; however, it has the highest reconstruction loss, emphasizing the importance of our semi-supervised training approach.

Table 4.5: Results of our ablation study. Arrows next to the metrics indicate the direction of better performance.

	Motion Rec.		Motion Syn.		Personality Recognition					
Models	**Rot↓**	**Pos↓**	**Div↑**	**FID↓**	**O↓**	**C↓**	**E↓**	**A↓**	**N↓**	**Avg↓**
full consistency	0.337	0.036	7.035	0.644	1.930	4.733	2.271	2.714	5.900	3.509
w/o personality consistency	0.387	0.045	8.097	0.715	**1.667**	**3.315**	**1.026**	**1.858**	**3.955**	**2.364**
w/o cycle consistency	0.370	0.040	9.450	0.516	2.387	4.616	2.861	2.427	5.358	3.530
w/o discriminator	0.356	0.038	8.594	1.213	1.877	**3.315**	2.392	2.601	5.442	3.125
supervised	0.622	0.077	7.353	1.114	2.541	3.496	1.526	2.785	4.768	3.023
P-GeNeMF	**0.327**	**0.035**	**10.208**	**0.456**	2.406	4.552	2.493	3.377	6.319	3.829

One interesting observation from the ablation study is that the generally high loss for the openness and neuroticism traits may be due to the solely motion-based nature of the data. Personality recognition results from Erkoc et al. [7] also show low performance for these traits. Additionally, in our labeling process, the standard deviation of the openness and neuroticism labels is high (see Figure 3.2), which might suggest that discerning these traits from motion alone is particularly challenging.

Chapter 5

Conclusion

This thesis explores and compares two methods for transferring personality traits in animation: the handcrafted Laban Movement Analysis-based approach and a data-driven deep learning model, P-GeNeMF. The results demonstrate that both methods have unique strengths. The handcrafted approach produces realistic animations, specifically when transforming traits between extreme levels. In contrast, the data-driven method, P-GeNeMF, shows greater flexibility and effectiveness in transferring a more comprehensive range of personality traits, albeit with occasional overshooting in trait moderation.

Through comprehensive user studies and evaluations, it becomes evident that data-driven approaches hold significant promise for scalable, automated personality modulation in animated characters, while handcrafted techniques remain valuable for achieving nuanced realism. As a potential future work, the handcrafted modulation tool could augment the dataset, thereby increasing its size and potentially enhancing the training of data-driven models. Additionally, improving the performance of the data-driven model could involve incorporating more labeled animations and adapting paired comparison-based labeling processes to reduce label variability. Moreover, state-of-the-art style transfer approaches designed for NeRFs, such as [49, 50], could be adapted for the animation domain to refine further and enhance the quality of the animations.

Bibliography

[1] S. Xia, C. Wang, J. Chai, and J. Hodgins, "Realtime style transfer for unlabeled heterogeneous human motion," *ACM Transactions on Graphics*, vol. 34, no. 4, 2015. Article No. 119, 10 pages.

[2] K. Aberman, Y. Weng, D. Lischinski, D. Cohen-Or, and B. Chen, "Unpaired motion style transfer from video to animation," *ACM Transactions on Graphics*, vol. 39, no. 4, 2020. Article No. 64, 12 pages.

[3] P. Li, K. Aberman, Z. Zhang, R. Hanocka, and O. Sorkine-Hornung, "GANimator: Neural motion synthesis from a single sequence," *ACM Transactions on Graphics*, vol. 41, no. 4, 2022. Article No. 138, 12 pages.

[4] D.-K. Jang, S. Park, and S.-H. Lee, "Motion puzzle: Arbitrary motion style transfer by body part," *ACM Transactions on Graphics*, vol. 41, no. 3, 2022. Article No. 33, 16 pages.

[5] F. Durupinar, M. Kapadia, S. Deutsch, M. Neff, and N. I. Badler, "PERFORM: Perceptual approach for adding OCEAN personality to human motion using Laban Movement Analysis," *ACM Transactions on Graphics*, vol. 36, no. 1, 2016. Article No. 6, 16 pages.

[6] S. Sonlu, U. Güdükbay, and F. Durupinar, "A conversational agent framework with multi-modal personality expression," *ACM Transactions on Graphics*, vol. 40, no. 1, 2021. Article no. 7, 16 pages.

[7] Z. Erkoç, S. Demirci, S. Sonlu, and U. Güdükbay, "Skeleton-based personality recognition using Laban Movement Analysis," in *Understanding Social*

Behavior in Dyadic and Small Group Interactions, Proceedings of Machine Learning Research, pp. 74–87, PMLR, 2022.

[8] D. Dotti, M. Popa, and S. Asteriadis, "Being the center of attention: A person-context cnn framework for personality recognition," *ACM Transactions on Interactive Intelligent Systems*, vol. 10, no. 3, 2020. Article No. 19, 20 pages.

[9] H. E. Cattell and A. D. Mead, "The sixteen personality factor questionnaire (16PF)," in *The SAGE Handbook of Personality Theory and Assessment*, vol. 2, pp. 135–159, Thousand Oaks, CA, USA: SAGE Publications, Inc., 2008.

[10] R. K. Hester and W. R. Brown, "Eysenck personality inventory: A normative study on an adult industrial population," *Journal of Clinical Psychology*, vol. 36, no. 4, pp. 937–939, 1980.

[11] K. C. Briggs, *Myers-Briggs type indicator*. Palo Alto, CA, USA: Consulting Psychologists Press, 1976.

[12] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of Personality*, vol. 60, no. 2, pp. 175–215, 1992.

[13] S. C. Guntuku, L. Qiu, S. Roy, W. Lin, and V. Jakhetiya, "Do others perceive you as you want them to? Modeling personality based on selfies," in *Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia*, pp. 21–26, 2015.

[14] J. Fu and H. Zhang, "Personality trait detection based on ASM localization and deep learning," *Scientific Programming*, vol. 2021, 2021. Article no. 5675917, 11 pages.

[15] H.-Y. Suen, K.-E. Hung, and C.-L. Lin, "TensorFlow-based automatic personality recognition used in asynchronous video interviews," *IEEE Access*, vol. 7, pp. 61018–61023, 2019.

[16] S. Song, S. Jaiswal, E. Sanchez, G. Tzimiropoulos, L. Shen, and M. Valstar, "Self-supervised learning of person-specific facial dynamics for automatic personality recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 178–195, 2021.

[17] H. Salam, O. Celiktutan, I. Hupont, H. Gunes, and M. Chetouani, "Fully automatic analysis of engagement and its relationship to personality in human-robot interactions," *IEEE Access*, vol. 5, pp. 705–721, 2016.

[18] C. Palmero, J. Selva, S. Smeureanu, J. Junior, C. Jacques, A. Clapés, A. Moseguí, Z. Zhang, D. Gallardo, G. Guilera, *et al.*, "Context-aware personality inference in dyadic scenarios: Introducing the UDIVA dataset," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, (Piscataway, NJ, USA), pp. 1–12, IEEE, 2021.

[19] F. Gürpınar, H. Kaya, and A. A. Salah, "Combining deep facial and ambient features for first impression estimation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, pp. 372–385, Springer, 2016.

[20] S. Aslan, U. Güdükbay, and H. Dibeklioğlu, "Multimodal assessment of apparent personality using feature attention and error consistency constraint," *Image and Vision Computing*, vol. 110, 2021. Article no. 104163, 9 pages.

[21] V. Ponce-López, B. Chen, M. Oliu, C. Corneanu, A. Clapés, I. Guyon, X. Baró, H. J. Escalante, and S. Escalera, "ChaLearn LAP 2016: First round challenge on first impressions-dataset and results," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, pp. 400–418, Springer, 2016.

[22] Z. Shao, S. Song, S. Jaiswal, L. Shen, M. Valstar, and H. Gunes, "Personality recognition by modelling person-specific cognitive processes using graph representation," in *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 357–366, 2021.

[23] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr, "A very brief measure of the Big-Five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, 2003.

[24] IBM, "IBM Watson API." `https://www.ibm.com/watson`, 2015. Accessed: 10 September 2024.

[25] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR '19, pp. 4401–4410, 2019.

[26] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, ICCV '19, (Piscataway, NJ, USA), pp. 4570–4580, IEEE, 2019.

[27] C. He, J. Saito, J. Zachary, H. Rushmeier, and Y. Zhou, "NeMF: Neural motion fields for kinematic animation," in *Advances in Neural Information Processing Systems*, vol. 35 of *NeurIPS '22*, (Red Hook, NY, USA), pp. 4244–4256, Curran Associates Inc., 2022.

[28] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[29] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR '19, (Piscataway, NJ, USA), pp. 5745–5753, IEEE, 2019.

[30] R. Villegas, J. Yang, D. Ceylan, and H. Lee, "Neural kinematic networks for unsupervised motion retargetting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR '18, (Piscataway, NJ, USA), pp. 8639–8648, IEEE, 2018.

[31] D. P. Kingma, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[32] K. Aberman, P. Li, D. Lischinski, O. Sorkine-Hornung, D. Cohen-Or, and B. Chen, "Skeleton-aware networks for deep motion retargeting," *ACM Transactions on Graphics*, vol. 39, no. 4, 2020. Article No. 62, 14 pages.

[33] X. Alameda-Pineda, J. Staiano, R. Subramanian, L. Batrinca, E. Ricci, B. Lepri, O. Lanz, and N. Sebe, "SALSA: A novel dataset for multimodal group behavior analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1707–1720, 2015.

[34] A. Cafaro, J. Wagner, T. Baur, S. Dermouche, M. Torres Torres, C. Pelachaud, E. André, and M. Valstar, "The NoXi database: multimodal recordings of mediated novice-expert interactions," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ICMI '17, pp. 350–359, 2017.

[35] J. Romero, D. Tzionas, and M. J. Black, "Embodied hands: Modeling and capturing hands and bodies together," *arXiv preprint arXiv:2201.02610*, 2022.

[36] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of motion capture as surface shapes," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, ICCV '19, (Piscataway, NJ, USA), pp. 5442–5451, IEEE, 2019.

[37] S. Ghorbani, Y. Ferstl, D. Holden, N. F. Troje, and M.-A. Carbonneau, "ZeroEGGS: Zero-shot example-based gesture generation from speech," in *Computer Graphics Forum*, vol. 42, pp. 206–216, Wiley Online Library, 2023.

[38] M. Kobayashi, C.-C. Liao, K. Inoue, S. Yojima, and M. Takahashi, "Motion capture dataset for practical use of AI-based motion editing and stylization," 2023.

[39] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pp. 851–866, New York, NY, USA: ACM, 2023.

[40] F. G. Harvey, M. Yurick, D. Nowrouzezahrai, and C. Pal, "Robust motion in-betweening," *ACM Transactions on Graphics*, vol. 39, no. 4, 2020. Article No. 60, 12 pages.

[41] I. Mason, S. Starke, and T. Komura, "Real-time style modelling of human locomotion via feature-wise transformations and local motion phases," *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, vol. 5, no. 1, 2022. Article No. 6, 18 pages.

[42] S. Sonlu, Y. Doğan, A. Ü. Ergüzen, M. E. Ünalan, S. Demirci, F. Durupinar, and U. Güdükbay, "Towards understanding personality expression via body motion," in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Workshop on Multi-modal Affective and Social Behavior Analysis and Synthesis in Extended Reality*, MASSXR '24, (Piscataway, NJ, USA), pp. 628–631, IEEE, 2024.

[43] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[44] Unity Technologies, "Unity 2019." `https://unity.com/`, 2019. Accessed: 10 September 2024.

[45] Blender Foundation, "Blender 3.6.5." `https://www.blender.org`, 2023. Accessed: 10 September 2024.

[46] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Advances in Neural Information Processing Systems*, vol. 30 of *NIPS '17*, (Red Hook, NY, USA), pp. 5769–5779, Curran Associates Inc., 2017.

[47] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, vol. 30 of *NIPS '17*, (Red Hook, NY, USA), pp. 6629–6640, Curran Associates Inc., 2017.

[48] C. Guo, X. Zuo, S. Wang, S. Zou, Q. Sun, A. Deng, M. Gong, and L. Cheng, "Action2Motion: conditioned generation of 3D human motions," in *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, pp. 2021–2029, 2020.

[49] K. Liu, F. Zhan, Y. Chen, J. Zhang, Y. Yu, A. El Saddik, S. Lu, and E. P. Xing, "StyleRF: zero-shot 3D style transfer of neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR '23, pp. 8338–8348, 2023.

[50] Y. Chen, Q. Yuan, Z. Li, Y. Liu, W. Wang, C. Xie, X. Wen, and Q. Yu, "UPST-NeRF: universal photorealistic style transfer of neural radiance fields for 3D scene," *IEEE Transactions on Visualization and Computer Graphics*, In press.

Appendix A

User Study

We provide the screenshots of each user study. Figure A.1 shows a screenshot from Study 1. The aim of this study is to demonstrate the effectiveness of our handcrafted modulation tool and data-driven method in influencing the personality of content animation when transferring personality traits from personality animation. The participants compare four different animations: the content animation, the result of the handcrafted method, the result of the data-driven method, and the personality animation. There exist twenty tasks and six questions for each task. The screenshot shows only the first question in the first task. Each task is assigned randomly to a participant.

Figure A.2 shows a screenshot from Study 2. The aim is to modify each personality trait in both positive and negative directions for content animation to evaluate the effectiveness of the handcrafted tool and data-driven approach. The participants compare three different animations: the content animation, the result of the handcrafted method, and the result of the data-driven method. Again, there exist twenty tasks and six questions for each task. The screenshot shows only the fourth question in task one. Each task is assigned randomly to a participant, as in Study 1.

Figure A.3 depicts a screenshot from Study 3. This study focused on the

content, style, and personality resemblance of character animations. The participants compare three different animations: the content animation, the result of the data-driven method (the center animation), and the personality animation. There exist twenty tasks and three questions for each task. The screenshot shows the first task. Like in other studies, each task is assigned randomly to a participant.

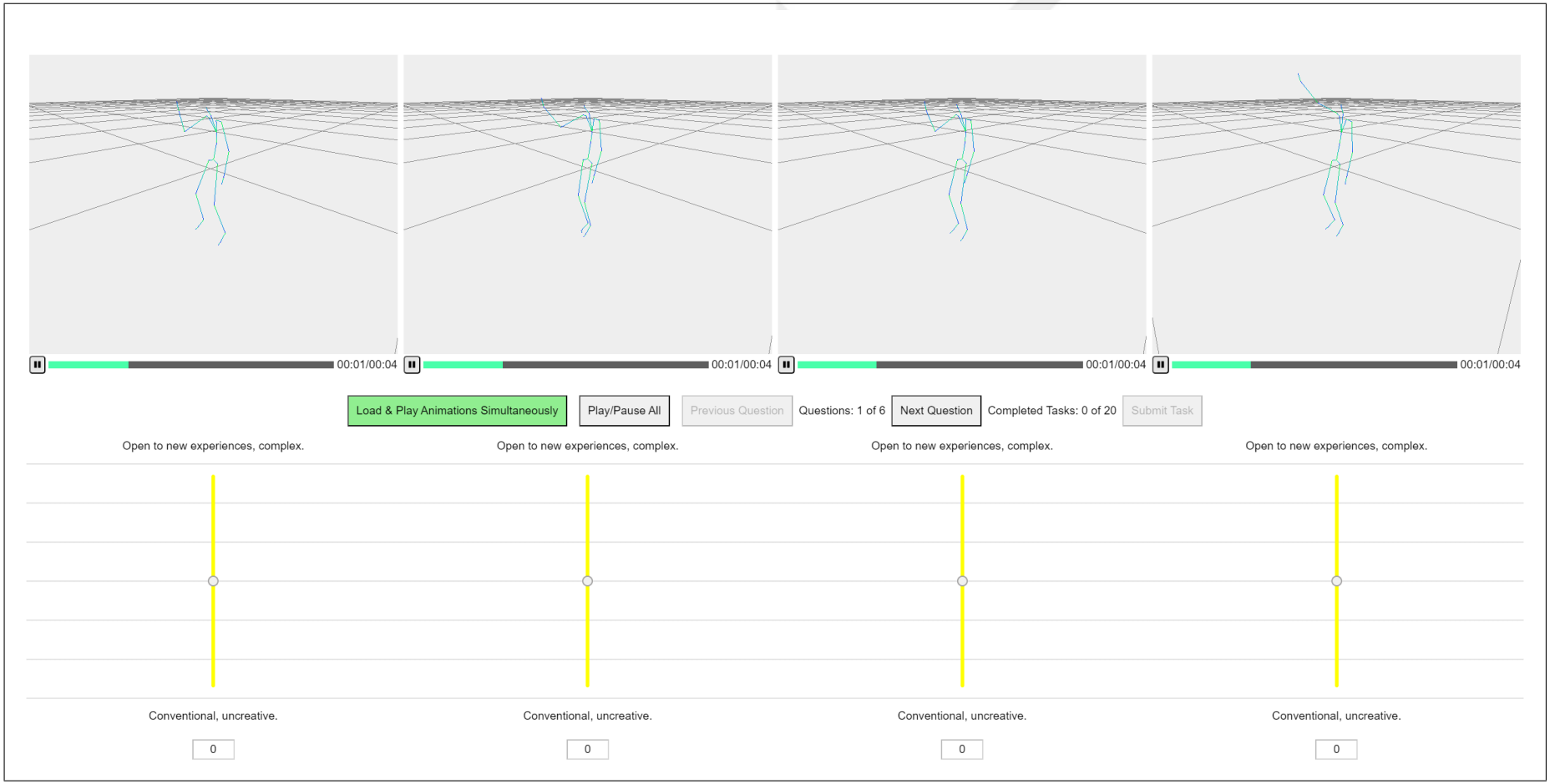

Figure A.1: The screenshot of Study 1. The current view shows the first task with a question related to Openness.

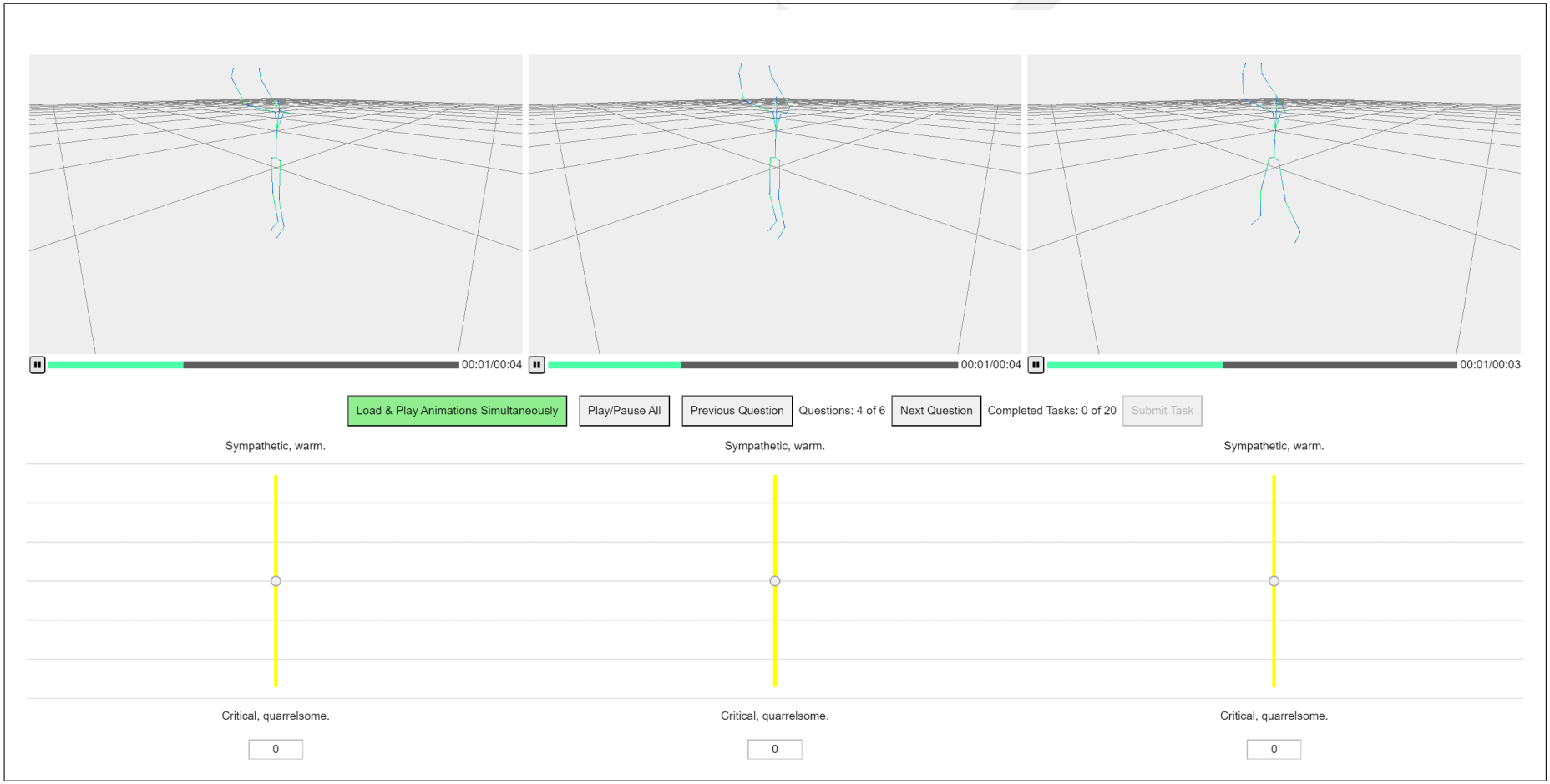

Figure A.2: The screenshot of Study 2. The current view shows the first task with a question related to Agreeableness.

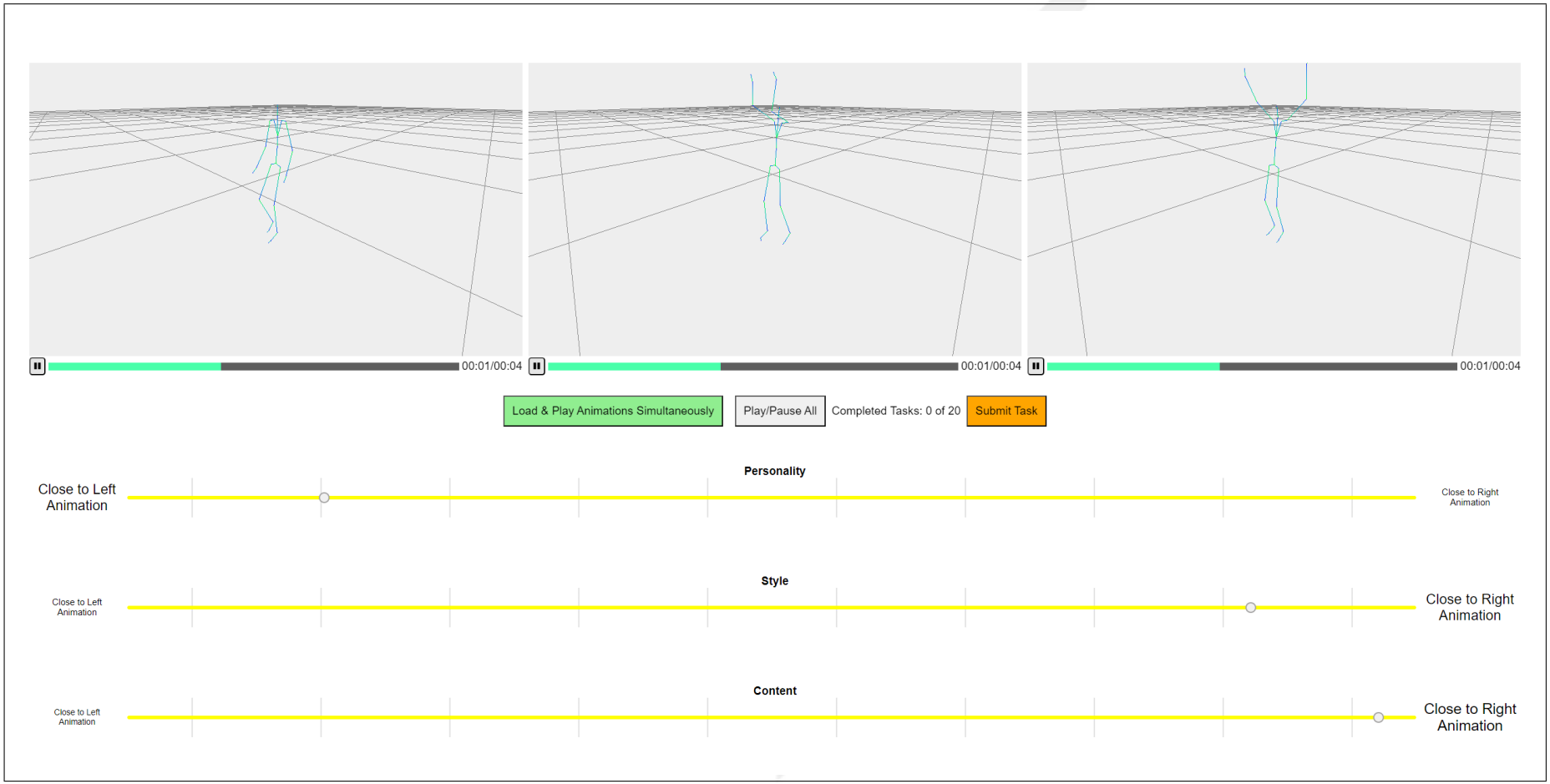

Figure A.3: The screenshot of Study 3. It displays the first task, which includes three questions: personality, style, and content resemblance. These questions are designed to compare animations generated by the data-driven model to those produced by personality-based and content-based approaches.