

TARGET-SENSITIVE INTERPERSONAL ANIMATION GENERATION WITH PERSONALITY EXPRESSION

A DISSERTATION SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

By
Sinan Sonlu
September 2025

TARGET-SENSITIVE INTERPERSONAL ANIMATION GENERA-
TION WITH PERSONALITY EXPRESSION

By Sinan Sonlu

September 2025

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

Uğur Güdükbay(Advisor)

Özgür Ulusoy

Yusuf Sahillioğlu

Ramazan Gökberk Cinbiş

Hamdi Dibeklioglu

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

TARGET-SENSITIVE INTERPERSONAL ANIMATION GENERATION WITH PERSONALITY EXPRESSION

Sinan Sonlu

Ph.D. in Computer Engineering

Advisor: Uğur Güdükbay

September 2025

Expressing personalities is essential in human animation for realism and improved communication. We focus on interpersonal animations that portray target-sensitive actions, such as handshaking, to build a data-driven model for personality-based dyadic motion synthesis. However, since datasets with personality annotations are rare, we first introduce personality labels to a subset of animations from well-known datasets with high style variance. To this end, we perform a user study for participants to rate the animation samples regarding perceived personality. We form a set of motion parameters based on various distances, angles, and volumes calculated over different combinations of body joints. Then, we identify the motion parameters that influence personality expression through an in-depth correlation analysis. These motion parameters influence the personality expression in our generative model. We build a transformer-based generative autoencoder that inputs the previous and current poses of a secondary agent, together with the previous pose and the desired personality of the main agent, to output the current pose of the main agent expressing the desired traits. We train this system using dyadic motion samples to learn the semantic actions, also utilizing a randomly selected personality sample for learning to replicate different personality styles based on their motion parameters. We evaluate the generative model by synthesizing motions where the main agent expresses different personality traits. The results suggest that our approach can generate a significant effect on perceived extraversion and agreeableness, while the influence on the remaining factors is limited, likely due to the secondary agent’s motion being a limiting factor. Consequently, the resulting model can effectively generate personality-specific poses in time but has less variety based on motion speed.

Keywords: Personality expression, interpersonal animation generation, Five Factor Model, Laban Movement Analysis.

ÖZET

KİŞİLİK İFADELİ HEDEFE DUYARLI KİŞİLERARASI ANIMASYON ÜRETİMİ

Sinan Sonlu

Bilgisayar Mühendisliği, Doktora

Tez Danışmanı: Uğur Güdükbay

Eylül 2025

Kişilik ifadesi, insan animasyonunda gerçekçilik ve pekiştirilmiş iletişim için gereklidir. Kişiliğe dayalı ikili hareket sentezi için veri odaklı bir model oluşturmak amacıyla, el sıkışma gibi hedef odaklı eylemleri tasvir eden kişiler arası animasyonlara odaklanılmıştır. Kişilik etiketleri içeren veri kümeleri nadir olduğundan, öncelikle, yüksek stil çeşitliliğine sahip, iyi bilinen veri kümelerinden animasyonlara kişilik etiketleri eklenmiştir. Bu amaçla, katılımcıların animasyon örneklerini algılanan kişilik açısından değerlendirdikleri bir kullanıcı çalışması düzenlenmiştir. Farklı vücut eklemi kombinasyonları üzerinden hesaplanan çeşitli mesafelere, açılara ve hacimlere dayalı hareket parametrelerinden bir küme oluşturulmuştur. Ardından, derinlemesine bir korelasyon analizi ile kişilik ifadesini etkileyen hareket parametreleri belirlenmiştir. Bu hareket parametreleri, üretken modelin kişilik ifadesini etkilemektedir. İkincil bir karakterin önceki ve mevcut pozlarını, ana karakterin önceki pozunu ve istenen kişiliğiyle birlikte girdi olarak kullanan ve istenen özellikleri ifade edecek ana karakter pozunu çıktı olarak veren dönüştürücü tabanlı bir üretken model oluşturulmuştur. Bu sistem, anlamsal eylemleri öğrenmek için ikili hareket örnekleri kullanarak eğitilmiştir ve ayrıca, hareket parametrelerine göre farklı kişilik tiplerini taklit etmeyi öğrenmek amacıyla rastgele seçilen kişilik örneklerinden yararlanılmıştır. Üretken model, ana karakterin farklı kişilik özelliklerini ifade ettiği hareketler sentezlenerek değerlendirilmiştir. Sonuçlar, bu yaklaşımın algılanan dışadönüklük ve uyumluluk üzerinde önemli bir etki yaratabileceğini, ancak geri kalan faktörler üzerindeki etkisinin sınırlı olduğunu, bunun da muhtemelen ikincil karakterin hareketinin sınırlayıcı bir faktör olmasından kaynaklandığını göstermektedir. Sonuç olarak, ortaya çıkan model, kişiliğe özgü pozları zamana bağlı ve etkili biçimde üretebilir, ancak hareket hızı açısından düşük çeşitliliğe sahiptir.

Anahtar sözcükler: Kişilik ifadesi, kişilerarası animasyon üretimi, Beş Faktör Modeli, Laban Hareket Analizi.

Acknowledgement

I would like to express my gratitude to my advisor, Prof. Dr. Uğur Gdkbay, for his guidance and support throughout my research and my PhD studies.

I would also like to thank Prof. Dr. Hamdi Dibeklioglu and Prof. Dr. Yusuf Sahillioğlu for their valuable advice and for participating in my thesis monitoring committee. I also thank Prof. Dr. zgr Ulusoy and Prof. Dr. Ramazan Gkberk Cinbiş for being a part of my thesis jury.

Special thanks to all my friends whom I had the opportunity to work together in various parts of my research, including but not limited to Yalım Doğan, Serkan Demirci, Ziya Erkoç, Arçin lk Ergzen, Musa Ege nalan, Fatih Pehlivan, Halil zgr Demir, Mehmet Akif Şahin, Ayda Yurtoğlu, Mehmet Efe Sak, Kerem Varnalı, Hamila Mailee, and Khasmamad Shabanovi. I also thank my friends Mustafa Can Çavdar, Hasan Balcı, Ahmet Burak Yıldırım, Sinan Çavdar, and Mustafa Utku Aydoğdu for their support.

I thank my mother, father, and sister for their endless love and support throughout my life. None of this would have been possible without them.

This research is supported by The Scientific and Technological Research Council of Turkey (TBİTAK) under grant no. 122E123.

Contents

1	Introduction	1
1.1	Motion Categories	5
1.1.1	Self Interaction Motion	5
1.1.2	Object Interaction Motion	5
1.1.3	Human Interaction Motion	6
1.2	Target-sensitive Motion	7
1.3	Publications	9
1.4	Thesis Organization	11
2	Background and Related Work	12
2.1	Computer Human Animation	12
2.2	Animation Design	13
2.2.1	Handcrafted Animation	15
2.2.2	Procedural Animation	16
2.2.3	Motion Capture Animation	17
2.2.4	Data-driven Animation	18
2.3	Personality Expression	21
2.3.1	Traditional Methods	22
2.3.2	Data-driven Methods	24
2.4	Motion Datasets	26
2.4.1	Single Person Action	26
2.4.2	Dyadic Action	27
3	Personality in Conversational Agents	29
3.1	Personality in Co-speech Gestures	30

3.2	Personality in LLM-based Agents	38
4	Personality in Motion	47
4.1	Motion Parameters	49
4.2	Perception Study	52
4.2.1	ZeroEGGS Animation Segments	52
4.2.2	User Study Details	59
4.3	Filtering Indecisive Annotations	60
4.4	Personality Correlation Analysis	64
5	Target-Sensitive Interpersonal Animation Generation	78
5.1	The Aim	78
5.2	The Notation	82
5.3	The Losses	83
5.4	The Architecture	87
5.5	Evaluation	88
5.5.1	The Stimuli	88
5.5.2	The User Study	90
5.5.3	Results	91
6	Conclusions and Future Research Directions	99
	Bibliography	103

List of Figures

3.1	Co-speech Personality Experiment Screenshot	33
3.2	Screenshots of models in LLM-based conversational agent.	39
3.3	LMA-based motion adjustments in LLM agent.	41
3.4	Box plots of each variation’s BFI-2-XS and LOES-S scores.	43
3.5	Box plots of each variation using personality and model grouping.	44
4.1	The used body joints and motion parameter types	50
4.2	The motion segmentation tool used with the ZeroEGGS dataset	53
4.3	The effect of sampling rate for ZeroEGGS segments	56
4.4	Two ZeroEGGS segments with small distance	57
4.5	Unity WebGL application used in the user study	59
4.6	Bandai distributions that fail normality for all factors.	62
4.7	Bandai distributions that fail normality for only one factor.	63
5.1	High-level summary of the interpersonal motion synthesis model.	80
5.2	Data flow used in calculating the losses.	84
5.3	Sample generated animation.	89
5.4	Example screenshot from the evaluation user study.	92
5.5	Bar plots of the user answers focusing on single questions.	98
5.6	Bar plots of the user answers combining different motions.	98

List of Tables

3.1	Pairwise Comparison of the Positive Personality Models	34
3.2	Pairwise Comparison of the Negative Personality Models	35
3.3	Mean Differences for each Opposing Personality Sample.	37
3.4	Pearson correlation results of the LLM agent.	44
3.5	Theme analysis results of the LLM agent.	46
4.1	Accepted and rejected annotations.	64
4.2	Motion parameters for openness in the Bandai dataset.	66
4.3	Motion parameters for conscientiousness in the Bandai dataset.	67
4.4	Motion parameters for extraversion in the Bandai dataset.	68
4.5	Motion parameters for agreeableness in the Bandai dataset.	69
4.6	Motion parameters for neuroticism in the Bandai dataset.	70
4.7	Motion parameters for openness in the ZeroEGGS dataset.	71
4.8	Motion parameters for conscientiousness in the ZeroEGGS dataset.	72
4.9	Motion parameters for extraversion in the ZeroEGGS dataset.	73
4.10	Motion parameters for agreeableness in the ZeroEGGS dataset.	74
4.11	Motion parameters for neuroticism in the ZeroEGGS dataset.	75
4.12	Selected motion parameters for the neural model.	77
5.1	User study questions for evaluation.	90
5.2	Pairwise comparisons of the results.	95
5.3	The results with factors combined.	96

Chapter 1

Introduction

Living beings exhibit interrelated behavioral, cognitive, and emotional patterns due to genetic and environmental factors [1]. Personality examines these patterns regarding their causes and common aspects emerging in behaviors. While animals, too, express personality traits [2], humans have a much deeper and varied interaction of personality [3]. Nonverbal communication cues of humans differ from those of animals due to physiological differences; for example, certain animals use facial expressions resembling the human smile, but in primates, this expression is associated with fear instead of happiness [4]. This thesis focuses exclusively on human personality. Specifically, we focus on generating interpersonal animations expressing non-verbal personality cues.

Interpersonal motion includes human movements due to communication between at least two individuals. This category does not include actions due to object interaction, locomotion, or biological phenomena such as breathing and sleeping. Although these other action categories also include personality cues, they involve different parameters and rules and thus are out of the scope of this study. For example, the human walk is highly expressive, and various gait features can signal health conditions [5] or personality variations [6]; however, since walking involves an active contribution of the legs, the parameters that affect personality perception are very different from everyday social actions. Movement

features, such as the arm and leg swing, are much dominant in walking [7] in contrast to the stationary communicative actions where hand [8] and head [9] related features are more prominent. This work focuses on the actions that involve limited locomotion; the animation can involve a small walk to approach the target agent for handshaking, or can include minor changes in the root position during speech gestures, but any expressive group dynamics like proximity [10] are not considered among the motion features for personality expression.

Interpersonal actions may or may not involve physical contact between the individuals. The actions that do not involve contact primarily include gestures used in nonverbal communication. While there are gestures with symbolic meaning, most emerge as a way to accompany speech [11]. Co-speech gestures can add a layer of communication, expressing the individual’s feelings regarding speech. Momentary micro gestures could reveal deception [12] or emotional stress [13] during speech. The posture of the body and actions could signal emotions [14] and personality [15]; these two concepts are closely related to each other [16]: personality examines the long-term behavioral patterns, while emotions mainly arise as a reaction to the environment and have short-term impacts.

The actions that involve physical contact often have culturally established symbolic meaning, such as in a handshake or high-five. Other types of interpersonal actions with contact can involve a physical goal, such as to carry or push someone, or can aim to show affection, such as in a hugging action. It is worth noting that actions with contact can vary dramatically, and using a common approach to all different action categories is challenging. For example, to synthesize a hugging action between two agents, full body awareness [17] is required; the rigid bodies that represent the limbs of the agents should not have self or in-between collisions that negatively impact realism. However, the raw animation data only stores the skeletal configurations and does not include any information on body proportions. Generating such actions plausibly requires using body models for the agents, simulating the collision avoidance for the limbs [18] during training, which can create a bottleneck for the data-driven models.

This study limits the motions to target-sensitive interpersonal animations

where a single target influences the main agent’s movements. We aim to generate the animation of the main agent given the motion of a secondary agent; in this case, the secondary agent acts as a target for the generated motion to either orient towards or make contact. For example, generated co-speech gestures could target the secondary agent’s root position, or generated handshaking animations could target the secondary agent’s hand position in time. We exclude any multi-target actions, such as speech gestures with more than one listener, where the main agent’s focus needs occasional shifts [19]; generating such motion requires inputting the animations of more than one character, and training such a system necessitates appropriate multi-party interaction data.

Recording interpersonal actions with contact is challenging due to overlaps between the limbs of the interacting parties. Consequently, datasets focusing on physical contact between the characters are less common than those focusing on the general actions. As a result, we limit our physical contact actions to two categories: handshaking and high-fiving. For general actions without contact requisites, we follow an unpaired training approach. An idle listener can accompany any conversational action without a specific contact requirement; for this purpose, the listener is positioned within an acceptable distance to the main agent, aligned with the agent’s root direction. This behavior allows both actions with contact constraints and non-contact single-agent actions to utilize the same pipeline that inputs the secondary agent’s movement to drive the generated animation.

Since personality-annotated animation samples are scarce, we utilize single-agent animations during the training of the dyadic actions to benefit from the diversity of such datasets. Most datasets that focus on dyadic actions do not involve various movement styles, and they often focus on the action and the interpersonal constraints. For example, the InterGen [20] dataset focuses on generating dyadic motion from text and forming accurate contact points between the agents. Although the dataset has a good variety of interaction categories, it lacks any style variance; they do not have multiple takes of the same action by different actors or with different acting styles.

A different dataset, Inter-X [21], which we use for training our interpersonal

motion synthesis model, includes different takes of the same action by various subjects of different personalities. Still, since the animations are not acted out, the variance due to personality is low. Furthermore, they do not aim to cover the whole personality range; they only record the self-reported personalities of the subject; consequently, certain common personality types are better represented, while some rare combinations are lacking. As a result, we utilize the single-agent animation samples from the Bandai [22] and ZeroEGGS [23] datasets during training to introduce personality-based variance. However, since these datasets do not contain dyadic animations and differ in terms of action categories, we form our loss functions to learn only the motion parameters that influence the personality aspect of motion.

To learn the personality aspect of motion, we first investigate the movement features that determine the perceived personality. To this end, we form motion parameters based on various measurements that can be calculated over body poses. Our motion features include distances between particular joint pairs, angles due to consecutive joint trios, speed and acceleration of individual joints, and volumes of the Convex Hull of a subset of joint positions. We perform a user study to annotate Bandai and ZeroEGGS samples based on their perceived traits, following the Five-Factor Model (FFM) [24] of personality. Then, we analyze any correlation between our movement parameters and personality factors to develop a set of motion descriptors that highly influence personality perception. Our personality-based motion parameters offer a way to accurately estimate and replicate certain personality types without copying the motion content.

For example, in a waving action, one may desire to employ the overall personality style of a different animation, such as an extraverted walking motion, where the figure expresses an energetic movement style. Generating the extraverted waving action requires replicating certain aspects of the aforementioned walking animation; however, if we copy the action for each joint, we would lose the original semantic content. We aim to alter the input animation based only on its personality style without disturbing its semantic content. Our personality-based motion features help replicate the personality without copying the semantic motion. The minimality of the features prevents the resulting system from learning

to replicate animations based on their personality features; instead, the motion content is dictated by the input motion, and it is adjusted to express the desired personality by utilizing our motion parameters.

1.1 Motion Categories

Our interpersonal animation synthesis model generates the main agent’s motion based on the secondary agent’s motion. We group human motion into three categories based on the rules of interaction, which influence the input parameters of the data-driven solution:

1.1.1 Self Interaction Motion

This category includes motion that does not involve any other human or object. For example, sleeping or stretching actions do not need a secondary target to perform the related motion. The movements of the self may or may not include any self-contact; for example, a thinking motion with head-scratching would need an action target, but since the target is already included in the body of the self, no external information is required.

The generation of pose-based *Self Interaction Motion* requires inputting the target animation frame only; the network can learn the intertwined structure of the motion from the skeletal structure. For example, the jumping action starts with a specific configuration of the limbs dictated by the current time of the animation.

1.1.2 Object Interaction Motion

This category includes interaction with a non-human body, usually a non-living object that can be moved or transformed due to the motion. For example, actions

like eating or carrying require an external target object that changes state due to the action; the action is directed at the object’s position, and the object’s reaction is due to physics. The related object may not move or change after the action; for example, if the person is trying to push the wall and the wall is unaffected, such actions are still examined in this category as they require an external target. In this case, the external object does not move by itself or based on known physical rules, so it can be calculated based on a given environment state. We include locomotion in this category since the object’s state, in this case, the environment, is the external immobile body that the person interacts with.

Generating *Object Interaction Motion* requires inputting the object’s starting position and the physical qualities (size, shape, weight, friction, etc.) in addition to the target animation frame. The network could involve a physics simulation to determine the object’s position in the generated animation frame. For simplicity, the object can be attached to a limb of the body during movement; for example, while holding a glass, the motion is minimally affected by the object’s physical state and can be assumed to move together with the hand position. The network should output the current position and state of the object in addition to the generated body configuration. The object’s current state could involve a complex geometric output for objects that can deform.

1.1.3 Human Interaction Motion

Motions in this category include interaction with a human body that is not directly affected by the self-action. Movements such as handshaking or hugging involve an action from both parties, but the rules of the interaction are only partially dictated by physics. In this case, the main driving force is the social rules of the interaction, and the other person involved in the action does not need to follow a specific pattern. For example, while the self holds a hand out, the other person can contribute to the handshake, but can also reject it by performing a different action. The actions of this category may or may not involve contact. For example, a high-five action requires short contact between the hands of both

parties at a specific time; on the other hand, hand gestures during speech are directed at a specific target (the other person) but do not involve any contact.

Generating *Human Interaction Motion* can be considered in two different ways. We can generate the motion of both parties, where inputting the target animation frame would be enough; the animation time dictates the current pose of both bodies. The more interesting concept is when we generate the self-animation to accompany the other person’s actions. In this case, to generate the current pose of the self, we must input the current limb configuration of the other body. This configuration includes the joint positions and rotations of the other person. We also need the target animation frame since time can drive the action. Suppose a touch action where the other person stands still; the current pose of the other person is not enough to determine the generated pose of the self; we also need to know the current time of the motion.

This thesis focuses on the third category of motion, specifically, the conversational action during interpersonal interaction. Such motion includes actions of symbolic meaning in communication, like a handshake, high-five, or wave, and co-speech gestures displayed during conversations with or without a connotation. The talking motion could involve multiple targets, such as talking to an audience; this work only focuses on talking animations between two sides. Similarly, certain dyadic motions can apply to more than two individuals, which is out of the scope of this work. For example, a hug could involve more than two persons, which requires the generated motion to consider multiple constraints.

1.2 Target-sensitive Motion

We follow a use case where the generated motion should accompany an input animation. The movement of the other human (secondary agent) is predefined and will not change due to the generated motion. The generated self-motion (main agent) should satisfy the social rules of the desired interaction. Additionally, the generated motion should express the target personality traits within the limits

of the interaction. That is, without breaking the original content of motion. For example, a walking motion should not turn into a run to alter its personality; however, it could become a more energetic or sudden walk.

We identify the main agent’s motion as *target-sensitive*, since the generated movements should comply with existing motion in certain target positions. In other words, the joint positions of the secondary agent act as targets to interact with for the main agent; the generated motion should be sensitive to these targets, and it should follow the social rules determined by the training examples, such as the distance between the main agent’s and secondary agent’s hands during a handshake motion. Note that, rather than using the full volume of the agents’ bodies, we focus on a solution based on the skeletal animation to simplify the problem; therefore, any self or in-between collisions are ignored. A solution that considers the rigid bodies of the limbs should utilize volumetric bodies for the agents, which dramatically increases the number of parameters.

Given an input animation for the secondary agent, we list the desired qualities for the generated main agent motion as follows:

- The generated motion should perform the target action category. For example, if the input motion is performing a handshake action, the generated motion should also include movements related to a handshake. This behavior requires the generated motion to follow the interpersonal constraints of the input animation, which is a list of features describing the relative distances and angles between the limbs of both parties.
- The generated motion should express the target personality traits using non-verbal cues related to our motion parameters. Since we follow a data-driven approach, the available animation samples performing the same action slightly vary. Among the samples of the same action, specific animations express motion features that support the desired traits; the generated motion should resemble animations with the motion features corresponding to the desired personality traits. However, the expressivity of each motion category is different; specific actions with more restricted rules have

less room to express personality, and thus, different actions also vary in personality expression.

- The generated motion should appear natural and human-like. For example, there should not be any physically impossible poses, and the limbs' speed should resemble the motion capture samples of the available datasets.
- The generated motion should have a natural flow when interpolated, which requires us to input the previous pose of the main agent to synthesize the current pose. Consequently, the resulting system requires an initial pose for the main agent to start the sequence, where we utilize the first frame of the main agent's ground truth motion.

Since dyadic motion datasets are more limited than the generic motion datasets that include *Self Interaction Motion*, we also utilize such single-character animations where applicable. For example, co-speech gestures of monologues can be matched with idling or listening animations for the other character. Furthermore, the personality literature on nonverbal movement cues applies to all three categories of motion and, therefore, can be examined together. To this end, our neural model uses a personality loss to learn to replicate any motion sample in terms of its personality-based motion parameters, which helps capture the stylistic motion choices of various personality types.

1.3 Publications

The following articles and conference proceedings papers are published as part of this thesis study:

- *Skeleton-based personality recognition using Laban Movement Analysis* [25]: This work focuses on using motion parameters based on Laban Movement Analysis (LMA) to estimate personality in human motion clips, which was published in *Proceedings of Machine Learning Research* in 2022.

- *Personality expression in cartoon animal characters using Sasang typology* [26]: This work analyzes personality expression in animated cartoon animals utilizing a categorization based on biopsychological traits. It was published in *Computer Animation and Virtual Worlds* in 2023.
- *Emotional expression as a means of communicating virtual human personalities* [27]: This work analyzes the personality aspect of emotional poses and facial expressions in virtual characters, which was published in *Modeling Visual Aesthetics, Emotion, and Artistic Style* in 2023.
- *Towards understanding personality expression via body motion* [28]: This work focuses on deriving motion parameters that highly influence personality perception in full-body animation. This work was published in *IEEE Conference on Virtual Reality* in 2024.
- *Personality perception in human videos altered by motion transfer networks* [29]: This work investigates the effects of combining different appearances and motion on personality perception using motion transfer networks. This work was published in *Computers & Graphics* in 2024.
- *Human movement personality detection parameters* [30]: This work explores the effect of utilizing movement parameters for personality detection in human animation. This work was published in *Signal Processing and Communications Applications Conference* in 2024.
- *Personality expression using co-speech gesture* [31]: This work uses a data-driven approach to express different personality traits in co-speech gesture animations. This work was published in *ACM Transactions on Applied Perception* in 2025.
- *Effects of embodiment and personality in LLM-based conversational agents* [32]: This work examines the effects of embodiment and personality expression in conversational agents that respond to user queries utilizing Large Language Models. This work was published in *IEEE Conference on Virtual Reality and 3D User Interfaces* in 2025.

- *Talk with Socrates: Relation between perceived agent personality and user personality in LLM-based natural language dialogue using virtual reality* [33]: This work analyzes personality perception of Large Language Model-based agents in Virtual Reality. This work was published in *Computer Animation and Virtual Worlds* in 2025.

1.4 Thesis Organization

The rest of this thesis is organized as follows. We explain the required background information and cite the related studies in Chapter 2. We analyze the perceived personality traits regarding co-speech animations and embodiment in conversational agents in Chapter 3. We examine the motion parameters that highly correlate with personality features based on the annotations we obtain from our user study in Chapter 4. We give details of our target-sensitive interpersonal animation generation system with personality expression in Chapter 5, including the results of our evaluation. Chapter 6 concludes the thesis, reflecting on the findings and mentioning the possible future research directions.

Chapter 2

Background and Related Work

This chapter provides background on computer animation and animation design, including different approaches to crafting computer animation. We include references to studies that aim at expressive animation generation, with and without personality. We also mention a subset of the publicly available human motion datasets, including the ones used in this thesis study.

2.1 Computer Human Animation

Computer human animation handles human motion as a series of predefined poses interpolated in real-time using a hierarchical skeletal body [34]. Although the animation’s poses, or keyframes, can be designed manually, motion capture techniques usually ease this labor-intensive process [35]. Motion capture can utilize special suits for the actors to wear [36]. This equipment may use optical tracking or gyroscopic sensors to capture positional or rotational limb information. Recent approaches to motion capture use image data directly with deep neural networks [37]. One problem with the methods that use two-dimensional input data for extracting motion is the lack of depth information. The overlapping

limbs or any movement perpendicular to the camera axis is difficult to reconstruct. To this end, the use of depth cameras [38] or multiple camera setups [39] improves the results, in addition to the use of wearable sensors [40].

The captured motion is usually noisy and requires additional steps for cleaning [41]. Usually, filtering is applied to reduce jitters due to occasional errors in capturing [42]. Recent approaches utilize neural networks for such denoising tasks [43]. While this cleaning stage produces more fluent motion, it could also remove certain micro-gestures that emerge spontaneously in everyday movements [13]. Consequently, more accurate motion capture methods are always desired. The captured animations usually contain a keyframe at each frame; thus, the poses can be played one after the other without interpolation. Still, the resulting motion can benefit from interpolation when increasing the frame rate, typically 30 or 60 frames per second (fps).

The computer graphics community is interested in realistically representing human motion, allowing easy variation generation. While motion capture techniques allow fine control over motion design, they are still labor-intensive and do not allow an easy-to-edit workflow. For example, each keyframe needs to be changed in a captured walk cycle to change the motion of the arms to hold an object using both hands. The pose of the hands and the related joints can be copied from other motions, but this could cause the combined motion to look unnatural. An algorithmic approach is required for creating plausible motion.

2.2 Animation Design

Traditionally, human motion used heuristic formulas to determine the body pose at a given period [44]. For example, carefully inspecting the human walk cycle reveals the geometric nature of the limbs in action. Procedural motion generation techniques involve forming rules that result in specific animations; different algorithms are needed for generating motions of different actions. While procedural motion enables the change of parameters to create variation or emotion [45], the

resulting motion usually appears robotic due to its rule-based nature. The most significant disadvantage of the rule-based procedural motion is that it is difficult to adapt the spontaneous human movements into geometric equations.

Recent data-driven approaches learn the common aspects of human motion using statistical methods, deep networks, and many samples. In contrast to the earlier procedural techniques, generative networks can learn loose rules that enable synthesizing more human-like motion. Since the distributions learned are based on probabilities, these approaches can easily introduce variation to the synthesized motion. The models used differ based on the task. The most common human animation tasks include motion generation from textual descriptions, end effector positions, noise, or an input animation to accompany, style transfer to express a specific action, emotion, or personality, and motion editing, where an existing animation is adjusted to satisfy certain constraints. These tasks mostly favor video games and conversational agents but are also used in movie crowd simulations. For example, the crowd that fills the background in a movie could utilize emotion-based motion generation to fill the scene with appropriate digital humans. The movements can portray fear or panic in a dangerous situation, or the crowd could portray more relaxed animations in a peaceful scene.

Crowd simulations require considering the positional configuration of the agents, which can be based on different attributes such as human relationships [46], physical strength consumption [47], or panic level [48]. The view distance of the crowd agents determines how much detail is required in the agents' body movements. If the agents are viewed from far away, their positional arrangement is more prominent and requires more attention. The body movement and facial animations are crucial for agents viewed at a close angle. Especially for conversational agents, the plausibility of the body movements is essential as a lack of sufficient realism in digital characters produces an uncanny effect [49]. Expressing personality traits helps ease such uncanny effects and make the animated characters likable [50]. Due to this benefit, many works aim at expressing personality or emotions in animated characters to introduce realism and remedy uncanny artifacts [51].

The digital motion is represented as a series of interpolated keyframes, where each keyframe’s pose usually represents an extreme of the corresponding movement. Different interpolation methods allow for overshooting the keyframes, such as bouncing, but they are uncommon and mainly used for non-human objects. Consequently, given continuous human motion, algorithms can extract keyframes to reduce storage at the cost of fine details [52]. While reducing the number of keyframes up to a threshold does not disturb the animation quality, it is impossible to represent specific actions with a limited keyframe count; each movement needs at least two keyframes to define its start and end configurations. Using the bare minimum can result in robotic animation since the limbs only move considering functionality and not style.

Four main approaches are used in digital animation: Handcrafted, procedural, motion capture, and data-driven. While all techniques are still used and are suitable for specific scenarios, there is a shifting trend towards data-driven animation. However, it is important to note that data-driven methods require input samples, which are animations created using some other approach, generally using motion capture techniques.

2.2.1 Handcrafted Animation

Handcrafted animation requires a specialist to define the key poses of the animated objects in time by defining their position, rotation, and scale. While other modifiers, such as skew, can be applied to the objects, the aforementioned three are sufficient for most realistic movements.

The designers usually set the starting and ending poses first, then adjust the intermediate poses by adding new keyframes on top of the interpolated ones. Each limb can have separate keyframes. For example, if the head needs to move only during a specific period, then it is sufficient to add keyframes to the head for that part of the animation, saving space.

Since handcrafted animation uses sparse keyframes, it uses less storage. One

other advantage is that it enables complete control over the resulting animation. Consequently, handcrafted techniques are combined with the more recent approaches to fine-tune the output animations. The significant disadvantage of the handcrafted approach is that the posing is done through complicated, labor-intensive interfaces.

Some methods, such as Inverse Kinematics (IK), ease posing in handcrafted animation. Using IK, the designer does not need to consider each limb in action and can form plausible poses by defining the end effector positions of the limbs. IK aims to form a connected configuration for the chain of limbs; additional joint constraints can be defined for more realistic poses [53].

There are also attempts to create a physical representation of the articulated figure to make the posing process more intuitive [54]. However, even with such sophisticated tools, crafting a realistic animation requires a vast knowledge of human motion and physics. Designers often use visual references while designing an animation that replicates specific actions, and it is still a challenging process to choose which poses to include as keyframes. Consequently, animation research always pursues novel ways to animate.

2.2.2 Procedural Animation

Procedural animation uses the mathematical nature of human motion to derive algorithms that represent actions as a function of time. The actions that procedural animation techniques can reproduce are often based on repetitive rules such as walking or running [55]. The rule-based animation is generally more suitable for animal motion [56], where intricate details are not crucial.

People can easily perceive the uncanny features in human animation. While animal animations can produce an uncanny effect if not well handled [57], this effect is less prominent than human motion. This effect is likely proportional to the viewer’s familiarity with the animated entity; the higher the familiarity and expectation, the easier such uncanny effects are triggered [58]. Consequently,

procedural techniques that result in more robotic motion are more suitable for non-human characters and non-realistic animation [59].

Procedural animation also includes generating motion for non-living entities, such as those used in fluid and explosion simulations [60]. This type of animation is often based on physical rules and can be produced by calculating the state of each particle under the effect. Such simulations are evaluated based on how well they portray the corresponding phenomena, and their parametrization can generate stylistic variations.

Procedural techniques can also drive the high-level decision control of animated characters [61]. In this case, characters portray predefined animations, but the system rules dictate when and how they play each animation. For example, a crowd simulation could determine the actors' positions while a rule-based system controls which animation each character plays based on their personality and emotion [62]. Although early computer animation often used procedural techniques, the current use cases of such animation are more limited, as it is challenging to design appropriate algorithms for realistic motion that can be generalized. Procedural methods are often used to adjust existing motion samples in more recent studies [63, 64].

2.2.3 Motion Capture Animation

Motion capture involves recording the movements of actual humans to replicate the subjects' actions digitally. There are many different techniques used in motion capture [65]. The most common ones include wearing special clothing with movement sensors like accelerometers, gyroscopes, or tracking devices. The motion sensors record the changes in rotation and speed of certain locations on the body [66]. Then, this information is replicated on an articulated figure as animation. The optical motion capture techniques involve a similar approach. This time, one or multiple calibrated cameras record the subject's movements. Special markers are tracked within the recorded motion and are converted into 3D data using optical calculations.

More recent approaches use data-driven pose estimation for the recorded videos, eliminating the need for special equipment and markers [67]. A key disadvantage of the optical methods is that the subject’s limbs must be visible in the recorded image to produce accurate results. Consequently, multi-person interactions are difficult to capture using optical approaches. Moreover, if the recording is to be done in an environment with limitations such as low light conditions or narrow spaces, then sensor-based methods are preferred.

The most common technique to design animations in the current industry is to use some form of motion capture [68]. This approach allows motion designers to work with actors to control the resulting animations to a fine degree. Still, the captured motion may require additional steps to clean the motions to prepare it as a product [41]. Designers can manually introduce minor changes to the resulting motion; however, altering motion capture data greatly requires automation since it contains dense pose data.

2.2.4 Data-driven Animation

The most recent approach to animation design is to use neural generative models trained on previously captured motion samples. Note that deep models can also be trained using reinforcement learning [69], in which case the configurations that achieve the training goals can be discovered without data. However, producing human-like results almost always requires realistic samples to imitate. Otherwise, reinforcement learning can exploit certain relationships to generate physically correct but non-human-like animations that satisfy the reward function [70]. Utilizing imitation learning by giving examples that result in high reward helps achieve both human-like and physically correct solutions that offer variation [71]. For example, such agents can well respond to changes in terrain or environment and adapt with plausible behavior [72].

While reinforcement and imitation learning offer integration of physical constraints, training such models takes longer since the explored feature space is

much larger. In contrast, solutions that utilize Generative Adversarial Networks (GAN) [73] can converge faster and offer quicker inference. GAN-based motion synthesis networks enable action conditioning to generate the desired action based on semantic class labels [74]. Utilizing autoencoders for this task enables latent space adjustments so that the different motion categories can be blended or mixed to create more controlled variation [75]. Another popular data-driven animation generation task is speech-driven co-speech motion synthesis [76]. In this case, studies utilize motion recordings with paired speech and learn a mapping from speech signals to the corresponding body poses. Such studies are not only limited to body motion synthesis and can also generate the corresponding facial expressions and mouth animations [77] or head and eye movements [78].

The most recent solution to synthetic human motion is the diffusion-based models [79]. Such models often utilize transformers to iteratively denoise input features to produce smooth animations. Human motion diffusion models are mainly text-guided, but can include additional mechanisms for learning physics-based relations [80]. For example, physical constraints can help remedy floating, foot sliding, and ground penetration artifacts. Additionally, spatial constraints help guide the generated motion via an input trajectory or key frame locations [81]. One key disadvantage of the diffusion-based solutions is that they require longer inference times and many more training samples, in contrast to those that utilize GANs and autoencoders. Consequently, studies still utilize GAN-based models for human motion synthesis, which can produce comparable results to state-of-the-art diffusion models [82]. Unlike image generation tasks, human animation has a very small feature space, which could explain why less sophisticated neural models can produce successful results.

Certain studies focus on dyadic human motion, particularly generating partner animations to accompany a given motion. For example, a transformer-based solution can synthesize the pose of a digital dance partner given the user’s current pose in real time [83]. Not every dyadic motion synthesis task focuses on generating an accompanying animation; some studies aim to generate both sides of a dyadic motion given the action context [84]. While this task requires more parameters, it has fewer constraints and can generate highly varying output. In establishing

a personality variance, having flexibility in motion is essential. For example, in a simple waving motion, if there is a small variance between the different training samples, the network may not create enough distinction among various outputs, and expressing different traits within the existing variance may not be possible. When we input the main agent’s motion to generate an accompanying animation, we add further constraints that narrow the variation; thus, it becomes essential to utilize additional losses to ensure personality-based variation.

One advantage of utilizing the autoencoder architecture to represent motion is the possible manipulations in the latent space [85]. The latent space representation of motion can be transformed to generate stylistic variation. This approach is especially useful when using Variational Autoencoders (VAE) [86]. Transformer-based VAE models that generate human animation can be conditioned on action categories to represent diverse motion [87]. Such models are also useful for motion estimation and compression [88], and can learn multimodal representations for transforming input sequences [89]. Conditional VAEs also support intention-guided motion generation utilizing end effector positions for the agent to reach at specific key frames [90]. To this end, the network can input the initial pose, the desired key frame, and the target position to generate a pose for the desired animation frame that reaches towards the target position with smooth interpolation. One necessity with every data-driven motion generation system is for the output poses to connect fluently. To this end, most studies utilize losses that prevent the model from generating very different poses between consecutive frames. Although such loss results in smooth motion, it may also prevent the network from representing shaky motion, which could be required for certain personality traits.

Overall, data-driven animation synthesis is a recent field in computer animation studies, and applications evolve with novel datasets and generative neural models. Style, emotion, and personality-based labeling of the existing human motion samples help introduce controlled motion generation.

2.3 Personality Expression

Many studies examine the link between perceived personality and different animation features. While body motion is a strong candidate for expressing personality, the viewers can perceive personality traits from facial expressions, body shapes, voice, dialogue, clothing, and style. These attributes can be designed to express a particular personality style. This behavior is instrumental in scenarios where the agent avatar is generated from scratch. For example, an online customer service can use personality-driven avatar design to deliver a customized experience based on user needs. Users may find specific designs more approachable or easy to connect with based on their personality traits.

On the other hand, most animated films or video games use previously designed characters, and the rules of the business may not allow altering the characters' looks drastically. For example, a character may need to look like a highly agreeable person in one scene without changing body shape, voice, or clothing. In this case, the dynamic elements of the animation, the motion and facial expressions, can be considered for expressing the personality. The style of motion is especially crucial for personality-based animation. The character can perform the same action following different patterns to signal different cues. For example, a walking animation can vary in speed, step length, hand swing distance, and spine and neck curvature; these stylistic motion choices have a perceptual psychological effect. While certain cultural elements can contribute to the meaning of such stylistic variation, the overall effect is usually universal.

The universality of personality in motion could be due to its physical nature. For example, introverts are often less energetic, and any motion that expresses low energy is slower and is attracted by gravity. In contrast, an energetic and hence more extraverted motion appears quicker and with an upright posture. Studies often utilize a high-level understanding of motion to relate to personality cues. To this end, Laban Movement Analysis (LMA) [91] is a popular theory that inspects human motion. LMA contains different categories that define motion, among them, LMA Efforts has strong relationships to personality [63].

LMA Efforts consist of *Space*, *Weight*, *Time*, and *Flow*. *Space* describes the intention towards the environment surrounding the subject. It can be *Direct* (single-focused) or *Indirect* (multi-focused). *Weight* describes the attitude towards gravity; in this sense, the action either opposes the gravitational force to show *Light Weight*, or works with the force to represent *Strong Weight*. *Time* describes the continuous nature of the motion. It can be either *Sustained*, such as in movements where time-based changes are not a dominant element, or *Sudden*, with the urgency of the action being pronounced. There are other LMA parameters, such as *Shape Qualities*, that describe the direction of body movement as *Rising*, *Sinking*, *Spreading*, *Enclosing*, *Advancing*, or *Retreating*. These six directions can be used for attracting the hand gestures towards some point in space, which is effective for personality expression [64].

In addition to LMA-based parameterization of motion, studies also utilize measurements based on various joints, such as the distance between specific elements or the angle of certain joints [15]. It is worth noting that LMA does not offer a standard way of quantification; rather, it offers categories to classify motion systematically. To this end, studies use different quantifications based on verbal LMA descriptions [92, 63, 64], which offer similar benefits.

2.3.1 Traditional Methods

The traditional methods for personality expression often utilize expressive poses or facial expressions that relate to certain personality traits [93]. A common workflow utilizes a perception study to assess the perceived traits in captured stimuli, which is analyzed for the common parameters influencing personality [94]. With a set of candidate cues for the corresponding personality types, studies often combine these attributes to develop rule-based adjustments for each personality factor. Even in simple motion cues such as the moving speed or acceleration of point-like objects, people can observe personality-related content [95]. When combined in realistic agent motion, the sum of cues affects the personality perception of the users. People often deduce personality from expressive motion cues

with high accuracy [96]. In the case of digital characters, this results in people similarly interpreting the expressive cues to reach a consensus about the agent’s personality. Unlike actual people, the digital characters do not possess ground truth personality values and are hence only assigned to perceptual annotations.

The hand motion is essential for accurate personality expression, as they have a dominant role in gesturing during communication [8]. The lack of information in short and simplistic motion sequences can cause the users to perceive certain personality factors together [15]. For example, if the subject lacks any clear cues for the intellectual aspect of the personality, the related factor openness can be perceived together with extraversion, resulting in a meta-factor, plasticity. The remaining three factors are combined under stability, which captures the individual’s emotional, social, and motivational stability. In addition to the magnitude of the facial expressions, the rendering style of the character is also important for accurate personality expression [97]. Cartoon characters can utilize more exaggerated poses and facial expressions in contrast to realistically rendered human characters, which affects the extent of expressive animations.

LMA-based movement modifications enable transforming personality expression in animation [63]. However, not all personality factors are affected similarly by such movement adjustments. The body movements influence the perceived extraversion the most; this factor is also the most common dimension shared by different personality theories. When combined with personality-based dialogue, speech, and facial expressions, agreeableness is also one of the factors that can be distinguished [64]. On the other hand, more intellectual factors, such as openness, are often difficult to perceive in short animated sequences and therefore require more language-level interaction with the agent. Expressive virtual agents can utilize audiovisual sensors to track the user’s current status and attitude. They can respond to the changes in the user’s gaze and attention to support a natural conversation [98].

Individual low-level communication elements are often grouped utilizing a middle layer, such as LMA, to introduce a controlled variance. For example, *Indirect Space* changes for adjusting a motion can include increasing the hand distance in

the horizontal axis, rotating the feet to have a wider resting angle, and shifting the gaze of the agent among multiple targets [64]. Since such cues are usually small elements, they are not tested in isolation; the effect of the middle layer is evaluated utilizing all the low-level elements that contribute to the same target. On the other hand, individual micro gestures also contribute to personality expression and can offer a way to analyze stress [13]. There are also high-level elements that contribute to the personality. Rather than utilizing body motion or facial expression cues, such high-level features affect the overall behavior of the digital characters [99, 100]. For example, instead of how a character walks or speaks, the high-level control of personality-based behavior influences where the character walks and to whom the character speaks. An introverted agent may avoid speaking with other characters completely, while an extraverted one seeks social interaction. Such goal-based planning is more applicable to games and crowd simulations. Conversely, personality expression through audiovisual cues controlled by middle-level features can benefit all agent-driven applications.

2.3.2 Data-driven Methods

There is no complete solution to generate data-driven digital characters with personality expression. Instead, some systems can generate parts of an agent-based system with personality variation. For example, the dialogue text of a conversational character can be customized to express certain personality traits [101]. LLM-based solutions can induce personality in free-form dialogue [102]. The agent’s dialogue can be vocalized using automatic speech synthesis customized for expressing different personality types [103]. Such systems utilize vocal features like pitch, intensity, and speaking rate, as well as certain speech features such as length and frequency of pauses, as personality-related cues. The face of the agent model can be generated to express certain personality traits [104]. Although utilizing the physical appearance to communicate different personalities is not possible for existing characters, such agents can still use cues related to clothing or accessories [105]. For example, an agent wearing formal clothes can signal high conscientiousness.

While there are many data-driven solutions to generative animation, personality-based variation in such applications is lacking, likely due to the scarcity of personality-annotated human motion datasets. Studies utilize similarity to example poses for motion matching to deliver real-time character animations in human-agent interactions [106]. However, such applications lack any personality-based variety. Automatic personality assessment can benefit similar studies by delivering variation based on target personality types. Movement analysis helps personality estimation in data-driven systems [107]. To this end, studies utilize certain angle measurements due to different joint combinations with respect to time. Using multimodal features improves the accuracy in personality estimation in the crowd [108]. Features such as proximity between subjects, gaze, and movement trajectories help classify different personalities in crowd interaction scenarios. Gait features such as the extent of hand and foot movement in walking also reveal information about the subject’s personality [7].

In group interactions, head pose is an important feature for personality estimation [9]. Similarly, gaze behavior can signal personality and be classified based on certain patterns [109]. Facial appearance, body posture, and behavior are closely related to apparent personality and useful features in automatic personality assessment [110]. The increase of personality-annotated animation datasets can lead to advances in the automatic synthesis of personality-based motion. Consequently, conversational systems can utilize fully automated agents to express the desired traits through multimodal architectures. The current applications manually combine various subsystems to support expressive personality in embodied systems. For example, the agent’s dialogue is generated separately from speech and animations. Multimodal datasets can enable the generation of the complete behavior of the agent, which could lead to better compatibility between the different elements of the expressive digital characters.

2.4 Motion Datasets

Although there are data-driven approaches to extract 3D joint locations from video samples [111], the lack of depth information in 2D video samples results in poor motion quality. As a result, we focus on motion datasets where 3D joint rotations or positions are available. Among the datasets included, these 3D skeletal animations are either recorded using motion capture equipment or are extracted using multiple calibrated cameras with additional depth information, and therefore, in high quality as opposed to any motion extracted from single-camera videos.

There are certain datasets with only 2D data available, but worth mentioning as they include personality annotations. For instance, the AMIGOS [112] dataset includes 40 participants watching emotional video clips in single or group settings. Various physiological signal recordings, full-body videos, and personality profiles of the participants are available. In this data set, participants are not instructed to perform any actions and are mostly stationary throughout the samples. The SALSA [113] dataset contains two video recordings of an indoor social event with 18 participants. The individuals' head and body positions and orientations are annotated, not their body poses. For each participant, self-reported personality questionnaire results are given. In this data set, participants are not instructed to perform any special task, but appear to be moving around the given area. Multiple camera angles are available. Some dyadic actions (e.g., talking and pointing) are visible throughout the video but are hard to turn into joint positions automatically. We group the 3D motion datasets based on the number of subjects participating in the action.

2.4.1 Single Person Action

Datasets that focus on single-person motions are generally designed for action recognition research. These datasets often support various action types and styles, although they are mostly acted out. For example, the subjects are instructed

to perform specific actions with certain moods or emotions in mind. Since these datasets offer good variance, they are good candidates for introducing personality annotations using perceptual studies.

The MoVi [114] dataset includes 17 hours of video from 4 different points of view of 90 participants performing 21 actions, including categories like walking, stretching, pointing, waving, and throwing. The annotations contain the 3D poses of the actions. The Berkeley MHAD [115] dataset contains 11 action categories performed by 12 participants. Depth video from multiple angles and motion capture of the 3D joint positions are available for each sample. This dataset includes action categories like jumping, waving, clapping, and throwing.

The UTD-MHAD [116] dataset includes 27 actions performed by eight participants. A total of 861 video samples have 3D joint position annotations. Motion category examples include waving, clapping, pushing, walking, and catching.

2.4.2 Dyadic Action

Our target-sensitive motion generation system mainly focuses on dyadic actions; therefore, we rely on two-person interaction motions for the different action categories. Since many overlaps occur during dyadic actions, the datasets of this category have less accuracy in contrast to the single-agent animations. Although some multi-agent interaction datasets include dyadic actions as part of their multiple-subject sequences, we only focus on datasets where two subjects interact.

The ShakeFive2 [117] dataset includes 153 videos of 8 different interaction categories like fist bumping, handshaking, high-fiving, hugging, and explaining. The full-body skeletal annotation of the samples is also available. The SBU Kinect Interaction [118] dataset contains eight categories of dyadic interaction: Approaching, departing, pushing, kicking, punching, object exchange, hugging, and handshaking. Each category has at least 21 samples, and joint locations of the interactions are available.

The NTU RGB+D 120 [119] dataset has 114,480 video samples from 120 categories, where 26 of them are dyadic. Some dyadic interactions include punching, kicking, pushing, pointing, hugging, and following. 3D body joints of the video clips are available. The AIR-Act2Act [120] dataset contains video and 3D joint positions of 10 action classes from three different views. There are a total of 5,000 samples. Action categories include approaching, handshaking, high-fiving, calling, and threatening. The Talking With Hands 16.22M [121] dataset contains 50 hours of multimodal content of two-person, spontaneous conversations. 3D joint positions for the body and fingers are available. The videos also contain the dialogue of the conversations that can be associated with different emotional content.

The dataset by Aberman et al. [122] contains 605 motion capture animation segments in BVH format, acting in 16 styles, such as old, depressed, childlike, and angry. The ZeroEGGS [23] dataset contains 67 sequences of performed monologues in 19 different motion styles. The total duration of the animations is 135 minutes. The animations are provided as BVH files that also include the motion of the hand fingers. We manually segment the animations of ZeroEGGS into 3-7-second-long animation clips to obtain 1,291 animation samples for our study.

Chapter 3

Personality in Conversational Agents

Conversational agents assume the role of systems with a human-like interface that can use natural language to interact with users. Not every conversational agent is embodied; such systems can utilize text-only interaction or be based on voice only [123]. In our case, we are interested in embodied conversational agents and specifically the influence of stylistic agent motion and dialogue content on personality perception. Since conversational agents often establish one-to-one interaction with the user, this chapter focuses on single-agent scenarios where the user interacts with a conversational agent through a dialogue. The studies mentioned in this chapter are preliminary research for our main topic, personality in interpersonal animations.

Although the effect of LMA-based motion adjustments on personality perception is a well-studied area [63, 15, 64], how these methods combine with the data-driven co-speech gesture generation systems is relatively new. To this end, we explore the effect of combining stylistic co-speech gestures with LMA-based motion adjustment on perceived personality of conversational agents in a joint work with Halil Özgür Demir, published in ACM Transactions on Applied Perception [31].

Stylistic dialogue choices have a dominant effect on personality perception [101]. Although the effect of handcrafted personality-specific dialogue on the users’ perception has been previously studied [64], the effect of personality variation in agents using Large Language Models (LLM) is yet to be explored. LLMs can assume the role of different personalities through prompting [124]. We combine personality-based motion adjustments in embodied agents that answer user questions utilizing an LLM backend with specific personalities to explore the effect of assuming different personality traits on the user perception in conversational agents. Our study on personality in LLM-based agents is presented in the 2025 IEEE Conference on Virtual Reality and 3D User Interfaces [32], with the collaboration of Bennie Bendiksen and Funda Durupinar.

3.1 Personality in Co-speech Gestures

Gestures inherently contain personality cues [15]. One key element in gesture performance is the appropriateness of the hand movements to speech. The ZeroEGGS [23] system enables driving agent animations through speech using a data-driven solution. ZeroEGGS also supports zero-shot style control by motion examples; thus, the generated speech-driven gestures could mimic different emotional or personal samples. As the resulting output is an animation, it can be further adjusted for specific personalities using heuristics-based LMA mappings [64]. We examine the combined effect of co-speech gestures and LMA-based motion adjustments on personality perception, human-likeness, and motion appropriateness in our work published in ACM Transactions on Applied Perception [31].

We use the pretrained version of the co-speech generation system [23]. We feed the network with dialogue samples crafted for different personalities from earlier work [64] to obtain co-speech animations for each personality (high and low polarity of the five factors, resulting in a total of 10) of FFM using the neutral, happy, and sad styles. We then further adjust the neutral style animations to express the desired factors with LMA-based motion adjustments [64], and as a case without co-speech gestures, we use the same speech animation with LMA-based

motion adjustments only. We compare the efficacy of five different approaches to personality expression:

1. **Neutral Co-speech Gestures:** In this case, the speech-driven animations are generated using the neutral style. Since the speech animations used in training ZeroEGGS [23] contain appropriate movements to follow speech, we expect it to capture some of the personality-based cues in language. Hence, this approach should express the desired traits to a certain level, as the dialogues are crafted to fit the corresponding personality types. We use the abbreviation G for this model, standing for “gesturing.”
2. **Happy Co-speech Gestures:** This group of animations uses the speech-driven gestures with the happy animation style, which is portrayed with energetic and expanding hand movements. Due to this emotional bias, we expect this category to better represent high-energy traits such as high extraversion. Due to the active gesturing, these animations may not represent low extraversion or low neuroticism well. We use the abbreviation H for this approach, standing for the used “happy” style.
3. **Sad Co-speech Gestures:** In this group, we utilize the sad style with the speech-driven gestures, which result in less active and enclosing motion. With this group, we expect to better represent low extraversion, low neuroticism, and high agreeableness due to the subtlety of the gestures. The content is unaffected while the emotional style input alters the motion’s range and speed. In other words, all three styles perform similar gestures with a variance in their high-level styles. We use the abbreviation S for this one, standing for the “sad” style
4. **Neutral Co-speech Gestures with Personality:** This approach uses the neutral style for the generated speech-driven gestures and enhances them using our previous work that includes heuristics-based LMA adjustments [64] to improve personality perception in the desired direction. For example, for the high extraversion dialogue and the resulting co-speech animation, we apply the animation modifications to express high extraversion

to further improve the results. We expect the modifications to have a positive impact on most personality types. This version uses the abbreviation F, which stands for the “full” model.

5. **Static Gestures with Personality:** Since we also examine the effect of co-speech gestures on personality expression, we use the same heuristics-based animation modifications on a dialogue-invariant talking animation. We use the same gestures for all personality variations but modify them using our LMA-based system. We expect this version to have low motion appropriateness as the speech does not direct the body movement; however, we still expect some effect on personality perception due to motion adjustments. This model uses the abbreviation P, which stands for “personality-only” adjustments, without speech-driven gesturing.

We use the same agent model for all variations and utilize the Oculus Lip-sync [125] library for the mouth animations that follow the agent’s speech. The agent’s speech utilized the samples from earlier work [64], and was generated using IBM’s Watson API [126]. Integration of the ZeroEGGS samples to our conversational agent system is done by Halil Özgür Demir, while we focused on conceptualization and carrying out the user study.

We implement an online survey (see Figure 3.1) for our user study ¹ where users rate the samples that express the same personality using different approaches. We use the Ten-Item Personality Inventory (TIPI) [127] to assess the perceived personality traits. We also include questions on motion appropriateness and human-likeness. We utilize the crowd-sourcing platform, Prolific, to acquire online participants paid based on the time spent on the tasks; 57 users participate in the study with a mean age of 29.36 ± 9.09 .

For each personality factor $x \in \{O, C, E, A, N\}$, we have x^+ and x^- variants, resulting in 10 combinations. Each personality variation $x^\pm$ is expressed with each model $y \in \{G, H, S, F, P\}$; we compare the five models in each set, focusing

¹Bilkent University Ethical Committee for Human Research approved the user study with the decision number İAEK.2023.06.09.02.



Figure 3.1: Sample screenshot from the user study page that compares five approaches that aim to express the same personality type. The participants use the sliders below each video to rate the perceived traits, and use the buttons to navigate between different survey questions. This image was originally published in [31]. © Association for Computing Machinery.

on expressing $x^\pm$. For example, among the set that expresses E^+ , we compare the models $\{E_G^+, E_H^+, E_S^+, E_F^+, E_P^+\}$. We utilize Analysis of Variance (ANOVA) to compare the models of the same set to discover significant differences. We repeat the statistical test for each measured aspect to report the significant pairwise differences with Tukey Honestly Significant Difference (HSD) adjusted p-values. We report the significant findings for the positive polarities in Table 3.1, showing the significant differences between the models for OCEAN personality factors, as well as human-likeness (L) and motion appropriateness (M) measurements.

The models with personality-specific motion adjustments are rather successful in portraying high openness and extraversion, with Model P performing the best for these traits. The Model H performs the worst for high conscientiousness, likely due to its high-energy movements. Models F and P perform similarly well in expressing high agreeableness traits. Model S has significantly better human-likeness when performing positive neuroticism. Generally, the models with speech-driven co-speech gestures perform better in human likeness and motion

Set	Pair	O		C		E		A		N		L		M	
		Δ_O	ρ_O	Δ_C	ρ_C	Δ_E	ρ_E	Δ_A	ρ_A	Δ_N	ρ_N	Δ_L	ρ_L	Δ_M	ρ_M
O+	P – F	.236	.036	-.046	.980	.353	.001	.069	.926	.069	.937	-.004	.000	-.024	.999
	S – F	-.280	.007	-.039	.989	-.326	.002	-.258	.023	.103	.777	-.233	.124	-.264	.038
	P – G	.501	.000	.070	.908	.625	.000	.281	.010	-.073	.923	.253	.076	.163	.403
	H – G	.332	.001	.023	.999	.436	.000	.165	.303	.011	.000	.094	.873	.048	.986
	S – P	-.516	.000	.006	.000	-.678	.000	-.328	.001	.034	.996	-.229	.136	-.241	.074
	S – H	-.347	.000	.054	.964	-.490	.000	-.212	.096	-.051	.979	-.070	.953	-.126	.657
C+	P – F	.292	.007	.056	.954	.374	.001	.268	.010	.021	.999	.112	.798	.004	.000
	H – F	.240	.042	-.203	.077	.309	.011	.051	.972	.248	.033	-.288	.036	-.225	.114
	P – G	.328	.001	.091	.774	.410	.000	.261	.013	-.064	.944	.189	.334	.045	.989
	H – G	.277	.012	-.168	.209	.345	.003	.044	.983	.162	.323	-.212	.222	-.184	.282
	H – P	-.051	.975	-.259	.010	-.065	.961	-.217	.063	.226	.065	-.400	.001	-.229	.104
	S – P	-.425	.000	-.061	.939	-.498	.000	-.381	.000	.062	.950	-.256	.085	-.104	.798
E+	S – H	-.373	.000	.199	.089	-.433	.000	-.164	.266	-.164	.311	.145	.605	.125	.666
	G – F	-.318	.002	.032	.997	-.199	.190	-.121	.639	.123	.670	.059	.976	-.035	.996
	P – F	.194	.142	.010	.000	.312	.006	.129	.578	.098	.827	.174	.406	-.052	.981
	S – F	-.402	.000	-.051	.978	-.336	.002	-.213	.109	.133	.598	.078	.934	-.059	.969
	P – G	.512	.000	-.022	.999	.511	.000	.250	.037	-.025	.999	.115	.775	-.017	.000
	H – G	.384	.000	-.083	.884	.354	.001	.168	.312	-.003	.000	.056	.981	.047	.987
A+	S – P	-.595	.000	-.062	.958	-.649	.000	-.342	.001	.036	.995	-.096	.872	-.007	.000
	S – H	-.468	.000	-.000	.999	-.491	.000	-.260	.027	.013	.000	-.036	.996	-.071	.942
	P – F	.144	.492	-.147	.412	.311	.004	.132	.534	.066	.940	-.144	.599	-.047	.988
	H – F	-.046	.986	-.327	.001	.178	.253	-.110	.702	.346	.001	-.417	.000	-.299	.016
	P – G	.332	.002	.033	.995	.540	.000	.326	.002	-.047	.983	.050	.987	.121	.710
	H – G	.142	.510	-.146	.422	.407	.000	.084	.863	.233	.060	-.222	.174	-.131	.647
N+	S – P	-.407	.000	.066	.938	-.513	.000	-.237	.047	-.057	.965	.021	.000	-.080	.917
	S – H	-.217	.112	.245	.034	-.380	.000	.005	.000	-.337	.001	.293	.030	.171	.377
N+	S – H	-.023	.999	.270	.038	-.149	.459	.028	.997	-.221	.150	.335	.005	.235	.049

Table 3.1: Pairwise comparison of the positive personality sets, reporting mean difference (Δ_X) and Tukey HSD adjusted ρ values. Highlighted values indicate $\rho < 0.05$ with $\Delta > 0.5$, $0.5 > \Delta > 0.3$, and $0.3 > \Delta$. We report pairs with $\rho < 0.05$ and $\Delta > 0.3$ for at least one measurement type. The table was originally published in [31]. © Association for Computing Machinery.

appropriateness. However, they cannot represent the targeted personality traits compared to the models with personality-specific modifications. We observe that the models with a specific gesture style can improve the personality expression for similar traits. For example, using the happy style benefits high extraversion, and using the sad style benefits high neuroticism.

We report the pairwise comparison for the sets depicting the negative polarity traits in Table 3.2. We observe that the negative traits have less perception difference between our models. The happy style negatively affects low extraversion; the energetic movements of this style cannot represent introversion. Among the different factors, the most significant difference is negative extraversion, where Model F performs best. For negative agreeableness, the models perform very similarly. The models that aim for negative conscientiousness and neuroticism mainly influence the perceived openness and extraversion. Model F can express low neuroticism without disturbing human likeness and motion appropriateness.

Set	Pair	O		C		E		A		N		L		M	
		Δ_O	ρ_O	Δ_C	ρ_C	Δ_E	ρ_E	Δ_A	ρ_A	Δ_N	ρ_N	Δ_L	ρ_L	Δ_M	ρ_M
O-	S - P	-.228	.090	.149	.450	-.314	.002	-.130	.587	-.051	.979	.072	.955	.058	.970
C-	P - F	.390	.000	-.048	.981	.430	.000	.143	.441	.059	.977	.185	.304	.023	.999
	H - F	.271	.010	-.093	.820	.388	.000	.117	.638	-.014	.000	.077	.928	-.015	.000
	S - P	-.365	.000	.116	.667	-.335	.001	-.070	.922	-.197	.292	.010	.000	-.047	.985
E-	G - F	.312	.008	.183	.280	.254	.075	.247	.059	-.306	.012	.243	.131	.228	.156
	P - F	.310	.009	.009	.000	.299	.021	.051	.981	-.007	.000	.125	.742	.073	.949
	H - F	.499	.000	-.024	.999	.590	.000	.226	.105	-.120	.712	.032	.998	-.052	.986
	H - G	.188	.265	-.206	.171	.336	.006	-.022	.999	.186	.290	-.211	.249	-.279	.044
	S - H	-.283	.023	.198	.204	-.322	.010	-.051	.981	-.108	.785	.190	.353	.220	.182
N-	P - F	.354	.001	-.072	.928	.514	.000	.162	.319	.106	.742	.068	.965	-.018	.000
	H - F	.158	.425	-.208	.141	.355	.000	-.047	.982	.260	.026	-.222	.204	-.298	.006
	P - G	.448	.000	.016	.000	.596	.000	.341	.001	.012	.000	.135	.687	.045	.986
	H - G	.252	.050	-.120	.668	.438	.000	.132	.534	.166	.320	-.155	.566	-.235	.056
	S - P	-.522	.000	.035	.995	-.764	.000	-.431	.000	.055	.969	-.154	.570	-.196	.161
	S - H	-.326	.004	.171	.317	-.605	.000	-.222	.072	-.098	.793	.136	.683	.083	.873

Table 3.2: Pairwise comparison of the negative personality sets, reporting mean difference (Δ_X) and Tukey HSD adjusted ρ values. Highlighted values indicate $\rho < 0.05$ with $\Delta > 0.5$, $0.5 > \Delta > 0.3$, and $0.3 > \Delta$. We report pairs with $\rho < 0.05$ and $\Delta > 0.3$ for at least one measurement type. The table was originally published in [31]. © Association for Computing Machinery.

Utilizing the neutral co-speech gestures generally results in highly realistic, motion-appropriate animations. While some degree of personality expression is

achieved through the dialogue used in this alternative, it is rather limited. Using a specific style for speech-driven gestures improves the personality aspect when the targeted traits are compatible with the style used. For example, the happy style can better express high extraversion and high agreeableness traits, and the sad style can better express low extraversion. On the other hand, when the targeted personality type is not compatible with the used co-speech gesture style, the results are poor. For example, using the happy style dramatically harms low extraversion expression.

We observe that utilizing personality-specific motion modifications helps express many personality types; however, these modifications usually negatively affect human-likeness. Since the modifications can adjust the movements into less realistic pose configurations, Model P, where the input motion does not follow the speech, offers more control over the changes. The animation sample used for Model P has no dramatic movements; personality-specific modifications achieve the diversity. However, when the input to the personality-based movement modifications has inherent personality, further adjusting the personality can sometimes have mixed results. Consequently, we observe Model P performing better than Model F and Model G in expressing personality. On the other hand, due to utilizing the same animation, Model P has less motion appropriateness overall.

A different way to assess performance is through the difference in personality perception, which is done by utilizing the models of opposing polarities. For example, the magnitude of difference between E_F^+ and E_F^- can measure Model F’s success in expressing diverse traits. We report the differences for each measurement using the opposing personality samples of each model in Table 3.3. For successful models, we expect a high mean difference in the opposing polarity samples for the target factor, while there is a small difference in human-likeness and motion appropriateness.

Models F and P perform better in distinguishing opposing polarities for most factors, except for conscientiousness. All models perform poorly for conscientiousness, and we observe that the opposing polarity conscientiousness samples are perceived as more different in neuroticism. The agent’s personality is observed

	O+ - O-					C+ - C-					E+ - E-					A+ - A-					N+ - N-				
	F	G	P	H	S	F	G	P	H	S	F	G	P	H	S	F	G	P	H	S	F	G	P	H	S
Δ_O	.49	.21	.49	.34	.20	.15	-.01	.05	.12	-.01	.77	.14	.65	.33	.15	.51	.19	.41	.24	.21	-.33	-.20	-.52	-.37	-.07
ρ_O	.000	.019	.000	.000	.031	.071	.927	.539	.172	.928	.000	.149	.000	.000	.127	.000	.041	.000	.024	.027	.000	.038	.000	.000	.469
Δ_C	.28	.10	.29	.29	.14	.28	.21	.39	.17	.21	.22	.07	.22	.19	-.01	.48	.21	.35	.20	.28	-.47	-.39	-.55	-.42	-.32
ρ_C	.001	.237	.001	.001	.088	.000	.012	.000	.059	.008	.025	.416	.027	.045	.922	.000	.019	.000	.045	.002	.000	.000	.000	.000	.001
Δ_E	.49	.19	.56	.39	.20	.24	.06	.18	.16	.02	.76	.31	.77	.33	.16	.38	.07	.43	.32	.15	-.33	-.28	-.69	-.51	-.05
ρ_E	.000	.042	.000	.000	.029	.006	.477	.067	.117	.846	.000	.002	.000	.001	.131	.000	.452	.000	.001	.146	.000	.001	.000	.000	.554
Δ_A	.48	.24	.40	.36	.20	.11	.01	.23	.04	-.08	.42	.05	.50	.24	.04	.72	.40	.85	.62	.56	-.32	-.23	-.46	-.29	-.04
ρ_A	.000	.009	.000	.000	.032	.199	.927	.006	.644	.344	.000	.533	.000	.004	.696	.000	.000	.000	.000	.000	.000	.006	.000	.002	.648
Δ_N	-.38	-.15	-.24	-.23	-.16	-.47	-.22	-.51	-.21	-.25	-.42	.01	-.31	-.18	-.06	-.71	-.50	-.69	-.47	-.55	.63	.56	.60	.54	.42
ρ_N	.000	.085	.010	.019	.067	.000	.023	.000	.036	.010	.000	.908	.002	.073	.535	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Δ_L	.34	.03	.33	.16	.03	.20	-.02	.13	-.16	-.13	.21	.02	.25	.29	.06	.31	.05	.20	.08	.12	-.09	-.03	-.33	-.10	.09
ρ_L	.001	.797	.002	.115	.789	.041	.871	.192	.116	.151	.044	.822	.021	.004	.527	.001	.581	.063	.438	.253	.390	.775	.002	.331	.350
Δ_M	.33	.06	.28	.19	-.02	.16	.07	.14	-.05	.09	.31	.05	.18	.37	.08	.31	.17	.41	.20	.15	-.11	-.05	-.24	-.01	.14
ρ_M	.000	.553	.004	.042	.836	.062	.439	.117	.647	.316	.002	.598	.075	.000	.391	.001	.072	.000	.056	.138	.157	.549	.006	.920	.099

Table 3.3: Mean differences for each opposing personality sample. Δ_x shows the mean score difference, and ρ_x is the adjusted p-value for the significance of the difference. Columns depict the compared pairs, and sub-columns show different models. We highlight $p < 0.05$ with $\Delta > 0.6$, $0.6 > \Delta > 0.4$, $0.4 > \Delta > 0.2$, and $0.2 > \Delta$. The table was originally published in [31]. © Association for Computing Machinery.

in certain cases, and related traits influence perception [128]. This co-occurrence of similar traits is more common in conversational agents, especially if certain cues are lacking [64]. For example, we observe that the difference in extraversion also causes a difference in openness in models F, P, and H, likely because these traits often co-occur in life. In this case, although the system does not aim to express openness, the cues that express extraversion are perceived together with openness.

Generally, the perception of the personality factors is easier to distinguish when co-speech gestures are used. In contrast, the heuristics-based motion adjustments make the cues more pronounced, but consequently, they are more difficult to relate to a specific personality dimension. As a result, using co-speech gestures without additional motion adjustments is useful if we aim to influence only a specific personality trait, but with a small magnitude. Including personality-based motion adjustments improves the magnitude of the difference but also has implications for the perception of other personality factors. The lack of co-speech gestures negatively impacts motion appropriateness, especially for the agreeableness samples in Model P. Similarly, motion adjustments negatively impact human likeness and motion appropriateness in openness samples.

3.2 Personality in LLM-based Agents

LLMs offer highly accurate text completion [129], and when combined with user queries, they offer human-like responses. The chat completion feature of LLMs is utilized in conversational agents to provide realistic responses to natural language user inputs [130]. Based on the data used in training, LLMs utilize specific language features that users perceive as personality [131]. Personality questionnaires, which are originally used for assessing the personality of individuals, can be used with LLMs to estimate LLM personality [132]. Using various prompting techniques enables LLMs to use language cues that resemble specific personality types [102].

Although some studies explore the capabilities of LLM-based agents [133], the implications of personality-based motion adjustments on LLM agents are less explored. To this end, we implement an LLM-based conversational agent system with personality expression to investigate the effect of embodiment and personality expression. Our agent system uses previous work for personality-based motion modifications [64], and we adapt the system for more realistic agents generated by Reallusion’s Character Creator [134]. While the body joints are the same as those in previous Mixamo models, the use of facial shape keys differs in the new models, as they can have more intricate facial expressions. Rather than using Watson API [126] for speech synthesis, we utilize Windows’ local Text-to-Speech (TTS) solution for immediate response. However, running the resulting application requires a compatible Windows operating system. We utilize the GPT-3.5 Turbo model offered by OpenAI’s Chat Completions API for the LLM backend.

We implement a virtual tutor that answers the user’s questions on a subject of interest. The application begins by allowing the user to choose the conversation subject: quantum computing, blockchain technologies, transformer architectures, quantum mechanics, string theory, and general relativity. Upon choosing the subject, we display brief information on the study, and the user is presented with a randomly selected model (see Figure 3.2.

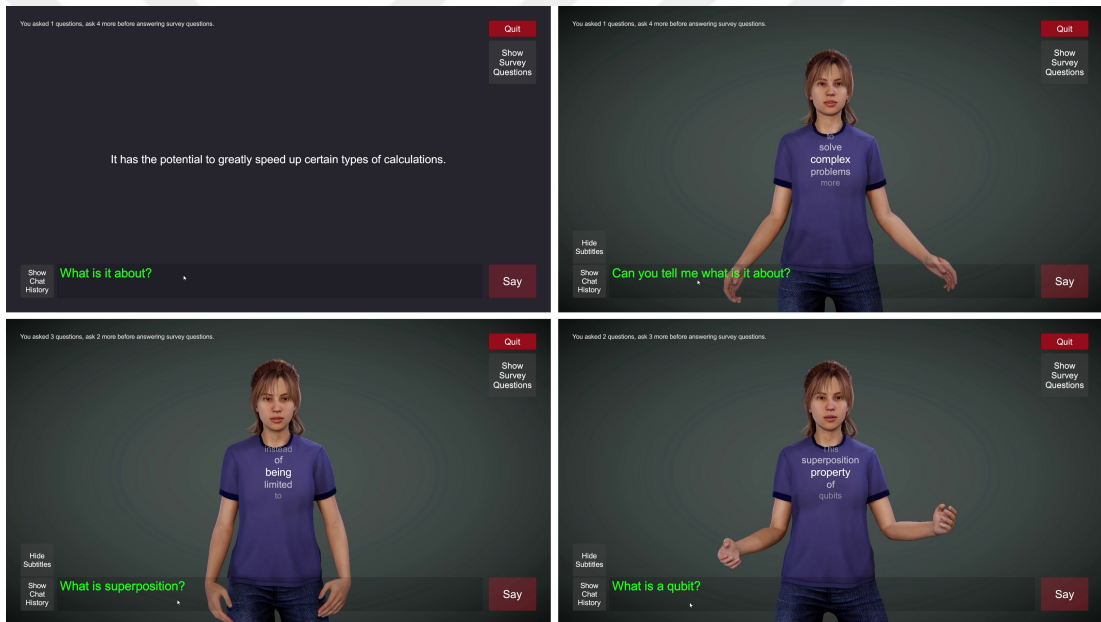


Figure 3.2: Screenshots of different models and variations: Model D (top left, dialogue only), Model A (top right, non-expressive, same animation for high and low traits), Model E-Low (bottom left, low-trait expressive agent), and Model E-High (bottom right, high-trait expressive agent). This image was originally published in [32]. Copyright © 2025, IEEE.

Our models differ in how they represent the conversational agent. Model D uses dialogue-based conversation only, without a visual representation for the agent. The agent’s responses are visualized as full-screen text, sentence by sentence, supported by the synthesized speech. In all models, the users type questions to the agent using the keyboard. The user can edit the input text before sending it to the agent. Clicking on the “Say” button turns off editing the input text, and the input text is sent to the API to generate the response.

We have two agent genders, male and female, that influence the agent’s voice and 3D representation, where applicable. We also support two ethnicities for each gender, resulting in four agent representations. Models A and E utilize a 3D model for the agent. In both models, the agent plays a randomly selected talking motion during their turn in the dialogue. We also utilize Oculus Lipsync [125] to animate the agent’s mouth based on the speech. Model A does not include personality-specific motion adjustments, while Model E utilizes LMA-based heuristic motion adjustments to express the desired traits.

We chose the combination of extraversion and agreeableness to test the system’s performance. We have a low personality and a high personality alternative with all models. The low personality variation aims to express the combination of low extraversion and low agreeableness, which appears as a less friendly and less social agent. In contrast, the high personality variant aims to express the combination of high extraversion and high agreeableness, which appears as a friendly and energetic agent. In all models, the LLM backend assumes the high or low personality variation using one of the following prompts:

- **High Personality:** “Act as an extraverted teacher teaching about *<topic>*, give friendly and polite answers.”
- **Low Personality:** “Act as an introverted teacher teaching about *<topic>*, give short and unfriendly answers.”

Models D and A express the personality through dialogue only. The animations of the agent in Model A do not include any personality-specific modifications.

On the other hand, Model E includes modifications to expressive movement. Based on the targeted personality polarity, the body movement and the facial expression of the 3D agent are adjusted to express the desired personality type. We summarize the motion adjustments for high and low personality variations in Figure 3.3.

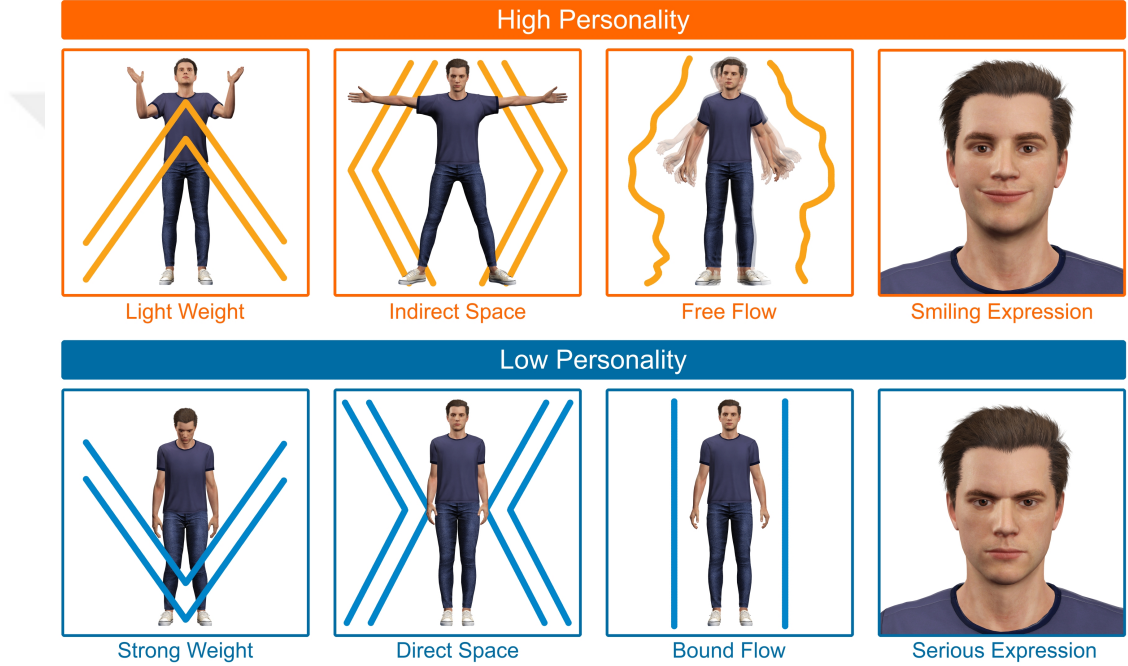


Figure 3.3: The summary of the LMA-based motion adjustments used for Model E. This image was originally published in [32]. Copyright © 2025, IEEE.

The high personality variant in Model E utilizes lightweight, elegant movements with an upright posture, indirect space corresponding to multi-focal attention, and free flow that results in an energetic, relaxed attitude. The agent’s high personality facial expressions are relaxed with occasional smiles. In contrast, the low personality variant uses strong weight, which results in a slanted posture, with direct space reflecting single-focused attention, and bound flow that shows strict movements. The low personality agent has a serious facial expression that shows frowns occasionally.

We implemented the LLM-based conversational agent system in Unity with the Windows build since we rely on Windows’ local speech synthesis library. We performed a user study with participants acquired from Prolific to test the effect

of personality and embodiment on perceived personality, learning, quality, and engagement outcomes. Two hundred ten unique subjects (99 female, 95 male, 16 not specified) participated in our study ², with an average age of 28.80 ± 8.57 . Each participant interacted with one variant of the system (Model D, A, or E, with either the high or low personality agent), and the gender and the 3D model of the agent were selected as random variables where applicable.

We use the extra-short form of the Big Five Inventory–2 (BFI-2-XS) [135] to assess the agent’s personality, which includes three questions for each factor. We use the Learning Object Evaluation Scale for Students (LOES-S) [136] to measure the conversation subject’s self-assessed learning, quality, and engagement. In this scale, learning questions measure the comprehension of the subject and the usefulness of the system for learning, the quality measures the ease of use, and engagement measures the user’s willingness to interact with the system. We also ask open-ended questions about participants’ understanding of the conversation subject, why they consider the subject important, if they learned anything new, if the system’s attitude influenced their experience, and whether they found the conversation interesting.

We used a 5-point Likert scale for users’ answers to the BFI-2-XS and LOES-S questions. We average the ratings that measure the same aspect to report the box plots of each model and personality combination in Figure 3.4.

We observe that all variations have highly positive perception quality and learning, while high personality variations received better engagement results. Users perceived all variations as highly conscientious, but the personality perception varies based on the expressed polarity for the remaining factors. The perception of high conscientiousness is likely due to the educative attitude of both high and low personality variations. Note that we use emotional stability as the reverse of neuroticism to report the results so that the high values correspond to traits with positive connotations for all personality factors.

²Bilkent University Ethical Committee for Human Research approved the user study with the decision number IAEK.2023.11.05.01.

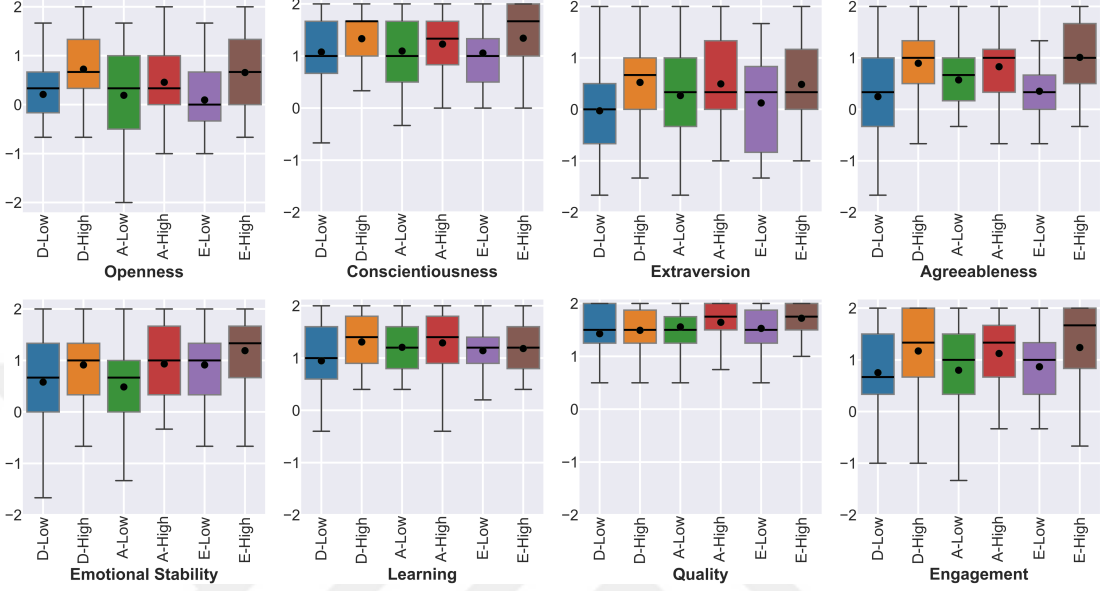


Figure 3.4: Box plots of each variation’s BFI-2-XS and LOES-S scores. This image was originally published in [32]. Copyright © 2025, IEEE.

We also report the BFI-2-XS and LOES-S score box plots using personality and model grouping in Figure 3.5, where we better observe the personality’s effect on perceived factors and engagement. The variation of the model has limited influence on any of the measured aspects, suggesting that embodiment does not affect learning, quality, and engagement. This behavior can be due to a novelty effect; the free interaction with the LLM-based agent could have diminished any effect due to embodiment.

We further investigate the Pearson correlation between the measured learning, quality, and engagement outcomes across all modes of the perceived personality factors in Table 3.4. In Model D, the LOES-S outcomes strongly correlate with the agent’s perceived conscientiousness. The lack of a visual representation for the agent could cause the learning outcomes to be associated with the perceived conscientiousness of the system. On the other hand, as we introduce a visual representation and expressive animation modifications, we observe a decrease in this correlation. Overall, correlations are the strongest in Model D and the weakest in Model E, utilizing embodiment. Personality-specific animations help distinguish the agent from the system’s performance regarding LOES-S measurements.

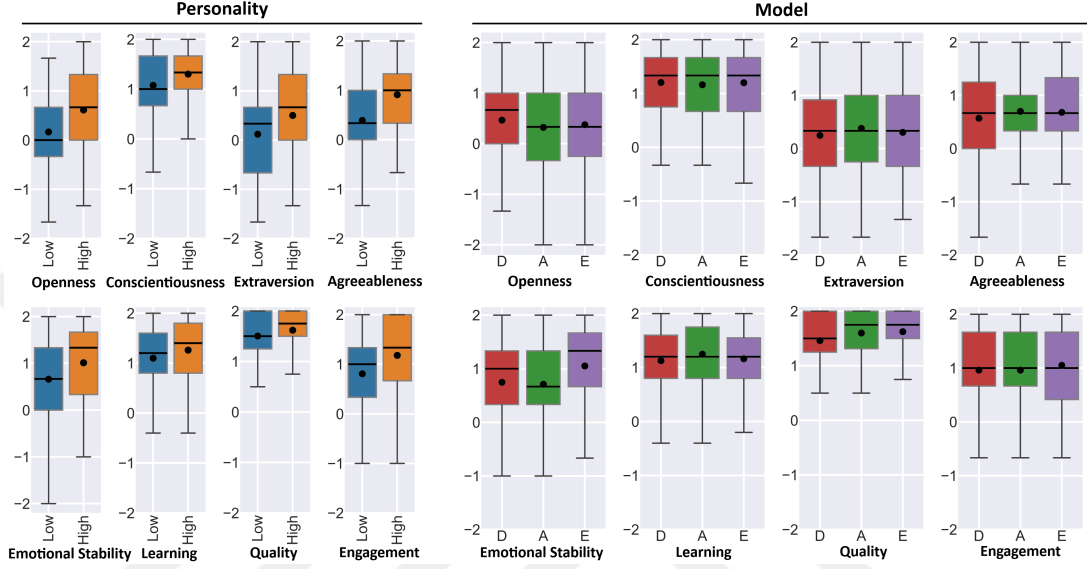


Figure 3.5: Box plots of BFI-2-XS and LOES-S scores assuming personality (Low-High) and model (D-A-E) grouping. This image was originally published in [32]. Copyright © 2025, IEEE.

Model	Cor.	O	C	E	A	ES
D	r_{Le}	.422**	.582**	.396**	.446**	.349*
	r_{Qu}	.437**	.654**	.279*	.325*	.190
	r_{En}	.417**	.608**	.287*	.422**	.109
A	r_{Le}	.351*	.348*	.456**	.398**	.208
	r_{Qu}	.215	.378*	.204	.150	.425**
	r_{En}	.504**	.332*	.505**	.541**	.233
E	r_{Le}	.266*	.405**	.281*	.184	.046
	r_{Qu}	.334*	.133	.056	.305*	.054
	r_{En}	.401**	.353*	.389**	.353*	.052

Table 3.4: Pearson correlation (r) between perceived personality and **L**earning, **Q**uality, and **E**ngagement. * indicates $p < .05$, ** indicates $p < .001$. The cell colors transition from weak to strong correlation. This table was originally published in [32]. Copyright © 2025, IEEE.

We also analyze the themes occurring in the participants’ answers to the open-ended questions. We mark the themes based on three categories:

1. **Benefit:** Whether the participant finds the system beneficial or not. The participants who find the system beneficial may comment on the detail and accuracy of the answers, the usefulness of the interaction, or the existence of an embodied animated agent. Those who find the system non-beneficial compare it to existing learning approaches, such as web search or watching tutorials, and find such interaction more practical than talking with a conversational agent.
2. **Learning:** Whether the participant indicates learning something new or not. The participants who learn new subjects related to the conversation often indicate that they did not know about the concept and comment on how the system influences their learning. The ones that indicate no learning often mention that they already knew about the conversation subject, and the interaction with the system did not introduce new knowledge, or even if it did, the extent of it was very limited.
3. **Interest:** Whether the participant finds the overall interaction interesting or not. The participants who find the system interesting mostly comment on the novelty of the experience and the advancement of technology. They find it interesting that the agent can accurately answer their questions with animations that enrich the experience. The participants who find the system not interesting mostly comment on the lack of realism of the agent’s voice and behavior, and mention that they are accustomed to similar systems.

We report the themes occurring in open-ended question answers for each variation in Table 3.5. We observe that, in general, the participants found the models with a 3D representation for the agent to be more beneficial. The participants who interacted with the high personality variant report better learning in Models D and A, in contrast to the low personality variation. Overall, the total number

of participants who find the system uninteresting is lower in Models A and E, suggesting that embodiment positively affects general success.

Variation	D-Low	D-High	A-Low	A-High	E-Low	E-High
Benefit	15	19	17	22	19	21
No Benefit	9	4	8	6	8	7
Learning	23	25	21	26	25	29
No Learning	6	6	5	2	7	2
Interesting	15	22	19	20	21	20
Not Interesting	3	2	0	1	2	1

Table 3.5: Theme analysis results. The numbers depict the occurrence of each theme for each variation. The cell colors indicate transition from low to high over 35 participant answers per variation. This table was originally published in [32]. Copyright © 2025, IEEE.

Chapter 4

Personality in Motion

Human motion naturally includes cues that express various personality traits. While there are theories that study the expressive nature of human motion, such as LMA, most of the literature on the relationship between body language and personality is limited to specific areas. For example, studies examine personality regarding posture and movement in the context of business negotiations [137] or seated interactions [138]. Comprehensive theories are lacking since body language can depend on social culture [139], which makes it difficult to combine similar movements with different meanings.

The same gesture could have a negative connotation in a different culture while having positive connotations in another [140]. Even the direction of the head shake that means “yes” or “no” is not universal; in Greece and Bulgaria, people shake their heads horizontally to indicate a “yes” gesture [141]. Such contradictions in the meaning of human gestures make forming a universally accepted theory difficult. A different reason is the versatility of human motion. The same action can be performed very differently under different constraints, altering the effect of individual limbs. For example, performing the same action while walking enables the legs to actively contribute to the expressive nature of the movement.

Waving hands while standing, leg distance, and foot angles can signal extraversion; however, such features may not be expressive if the subject is sitting. Furthermore, based on the object used for sitting, the features useful for personality estimation while standing could give contradicting results. For instance, a narrow seat could lead the subject to assume a closing posture regardless of personality. Consequently, we need subjects to act in a non-constrained environment without using culture-specific gestures to develop a generalizable mapping between motion and personality.

Not every co-speech gesture has a clear meaning; some gestures used in speech follow the rhythm of the speech or are there to emphasize the words spoken [142]. Movement-related features, such as speed or distance between certain body parts during interaction, can control the expressed personality of gestures [15]. Rather than focusing on communication elements with specific connotations like hand gestures, our goal is to examine the high-level features of general motion that influence personality.

It is also worth noting that personality can be assessed differently, including self-reported questionnaires. In our case, we are more interested in the apparent personality traits that the viewer perceives. For example, the actor used in the motion capture process can have a particular personality; however, due to the recording setup, the action performed, or the role of the subject, the personality apparent in the recorded motion can differ from the actor’s actual personality. The same subject can act differently to express different traits. Thus, we use perceived personality annotations collected through our user studies.

On the other hand, when perceptual annotations are lacking, the self-reported personality labels are also useful. Specifically, if the subjects are not instructed to act in different styles and the action happens naturally, we expect the self-reported personality traits to be mostly apparent [143]. Since the Inter-X [21] dataset with dyadic animations already contains self-reported personality annotations, and the number of samples in this dataset is too high for a budget user study, we use the available annotations. On the other hand, since the Bandai [22] and ZeroEGGS [23] datasets contain a good variety of different styles and offer more

manageable sample sizes, we conduct two user studies to introduce perceived personality annotations to these datasets.

Ultimately, we require a set of parameters useful for personality expression in general motion. For this purpose, we form potential movement features and analyze the importance of these features regarding personality expression. We aim to express personality in target-sensitive dyadic actions by utilizing a limited number of expressive motion features so that the synthetic animations replicate the desired personality style instead of the semantic action content.

When the whole motion is used for learning the personality, data-driven solutions can also learn undesired features. For example, a punching action is generally high in extraversion and low in conscientiousness; however, a system that aims to express these traits should not replicate a punching motion completely. Instead, it should capture the motion style of the punch to transfer it to different actions. Consequently, high-level motion descriptors correlating significantly with personality are essential for a generalizable solution for personality expression.

4.1 Motion Parameters

Forming parameters representing the whole action sequence is non-trivial since animation is time-variant. Each animation can include a different number of frames; as a result, any single frame parameter cannot represent the complete motion. On the other hand, frame-based parameters can be easily calculated using the positions and rotations of the human joints. Such geometric pose descriptors are common and are used in data-driven motion research [144]. These pose-based features can be used as time series data with recurrent or transformer networks [145]; however, such models usually require a limited feature size and do not perform well when the feature size is very high.

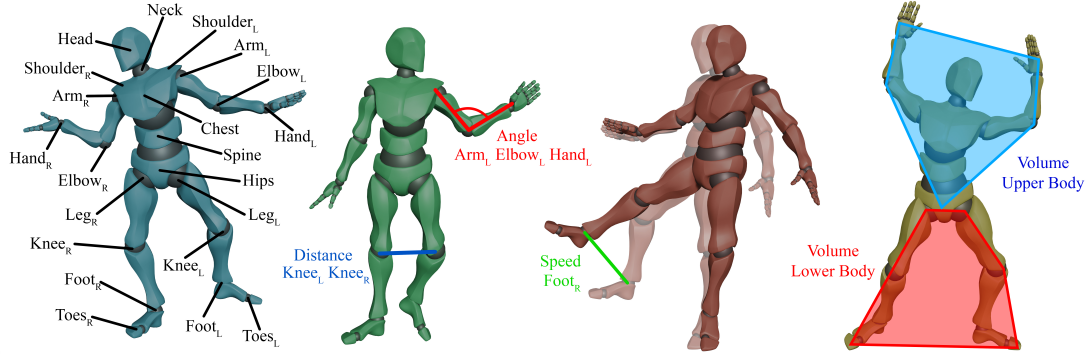


Figure 4.1: The body joints and motion parameters used in this thesis work. The used joints are the common subset of the three datasets we utilize. Since we focus on body motion, the finger joints are not included.

Using a data-driven approach, we want to determine the effective motion parameters in personality expression. Consequently, we form all the potential parameters and aim to analyze the correlation between each potential parameter and personality factor. As a result, using all the possible features with a neural model is infeasible for a feature importance analysis due to the high number of input features. Therefore, we use the average values of the pose-based parameters to represent the sequences. Such usage is common; for example, Aristidou et al. [92] utilize the minimum, maximum, mean, and standard deviation of pose-based LMA parameters to evaluate folk dance sequences. Inspired by their usage and to limit the number of variables that originate from a single parameter to one, we utilize the mean of our pose-based parameters across the animation frames to represent the sequences. For example, the average distance between the left and right hands becomes a motion parameter potentially affecting perceived personality.

Our motion parameters are based on five categories of measurements calculated using the 21 body joints common to the three motion datasets (see Figure 4.1). Distance-based parameters correspond to the Euclidean distance between two joint positions of the same frame. Angle-based parameters are calculated using the positive angle defined by the line segments that connect three joints in the same frame. Speed-based parameters are calculated as the Euclidean distance of

the same joint between two consecutive frames, and acceleration-based parameters are calculated using the difference between two consecutive speed parameters. Volume-based parameters are calculated as the volume of the Convex Hull of a subset of joints in the same frame. We calculate these parameters for all possible combinations for our personality correlation analysis.

Our five motion parameter types are calculated as follows:

1. **Distance** (D_{ab}): The Euclidean distance between the global positions of the joint a and joint b . The joint positions are calculated in the same frame of animation. This parameter type is calculated for all possible $\{a, b\}$ pairs; thus, there are ${}_{21}C_2 = 210$ combinations.
2. **Angle** (A_{abc}): The positive angle in radians that is defined by the line segments $\overline{ab}$ and $\overline{bc}$ for any combination of three joints. For any triplet $\{a, b, c\}$, the middle joint determines the angle, and the side joints a and c are interchangeable. We have 4410 such unique configurations using the 21 joints for the angle parameters.
3. **Speed** (S_a): The Euclidean distance between the positions of the same joint in two consecutive animation frames. Since this parameter requires two frames to calculate, there are $N - 1$ such values of a specific speed parameter for an N -frame animation. There are 21 such speed parameters.
4. **Acceleration** (C_a): The difference of the speed parameters in two consecutive frame pairs for the same joint. Suppose $p_a(i)$ is the 3D position of some joint a in frame i . If the speed of this joint in the frame i is $S_a(i) = |p_a(i) - p_a(i - 1)|$, then acceleration at the same frame is $C_a(i) = S_a(i) - S_a(i - 1)$. Since three frames are required to calculate acceleration, there are $N - 2$ such values of a specific acceleration parameter for an N -frame animation. There are 21 such acceleration parameters.
5. **Volume** (V_A): The volume of the Convex Hull defined by a subset of joints that contains at least four points. We do not include certain joints, like the chest or spine, that are mostly contained within the Convex Hull.

The candidate joints with their index numbers as used in V_A are as follows: *Head* : 4, *LowerArm_L* : 7, *Hand_L* : 8, *LowerArm_R* : 11, *Hand_R* : 12, *LowerLeg_L* : 14, *Foot_L* : 15, *LowerLeg_R* : 18, *Foot_R* : 19; the possible configurations with these nine joints correspond to 5,985 unique combinations.

4.2 Perception Study

Since the Bandai [22] and ZeroEGGS [23] datasets do not include personality annotations, we perform a perception study to introduce FFM labels for the samples using crowdsourcing. The Bandai dataset contains a good variety of short samples, each motion acted in 15 unique styles to introduce variety. The different styles are especially useful for invoking different personality effects. On the other hand, the ZeroEGGS dataset includes rather long (3-5 minutes) monologue sequences with various emotional categories. Using such long samples is difficult in a user study where participants often have a short attention span per task, so we first form short motion clips from the ZeroEGGS samples. Then, we run a user study using Prolific’s crowdsourcing platform with the selected samples from both datasets.

4.2.1 ZeroEGGS Animation Segments

To prepare shorter motion clips out of the ZeroEGGS samples, we implemented a WebGL [146] application using the ThreeJS [147] library that enables the user to create segments in BVH animation manually (see Figure 4.2). In contrast to existing animation editing software such as Blender, our tool enables easy preview of the segments marked. It supports editing the segments by clicking on the previously segmented areas indicated with different colors. Our tool also supports setting the beginning of a new segment just after the ending mark of a previous segment to easily form consecutive clips. During editing, we store each subsequence’s beginning and ending frames, which are later used in cropping the

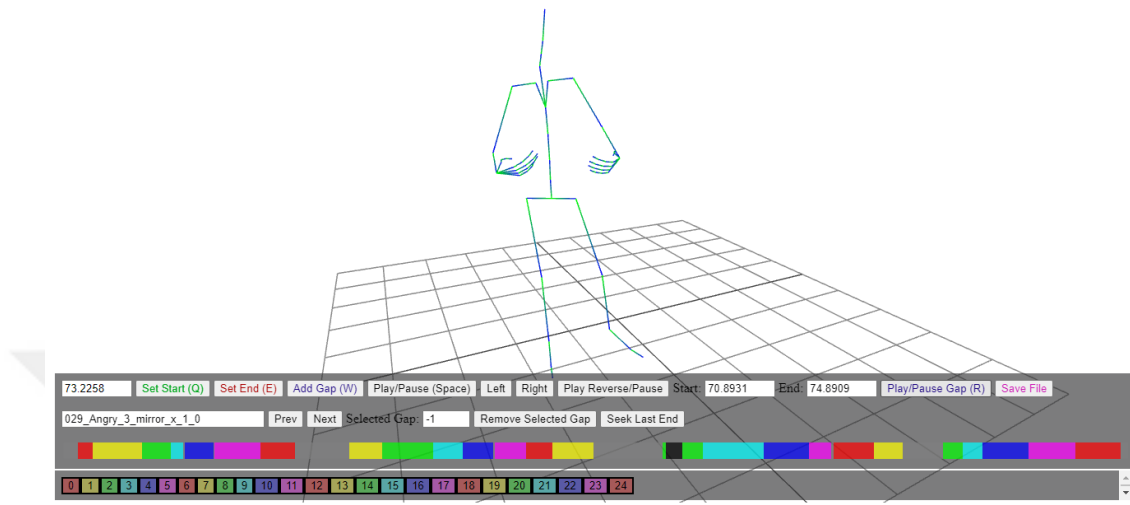


Figure 4.2: Our motion segmentation tool used in forming short motion clips using the long ZeroEGGS samples. The tool enables easy selection of different motion samples and supports setting each subsequence’s beginning and ending frames. The corresponding segments can be replayed and adjusted by clicking the ones indicated in different colors.

segments.

We manually segmented the ZeroEGGS animations into meaningful subsequences. During this segmentation process, we focus on choosing the beginning and ending frames that are relatively similar in terms of their pose so that the animations can be looped easily. We also pay attention to the gestures to finish their movement so that segments do not feel cut in the middle of the action. Thus, each segment finishes in an idle-like pose. We remove the idling parts of the animations so that each segment is a three-to-seven-second-long clip. We obtain 1,291 segments after this process. However, due to the limited budget, we selected a small subset of these 1,291 segments to be rated in our user study. Thus, multiple users can evaluate each sample, increasing the accuracy of the acquired personality annotations.

Since the ZeroEGGS dataset includes improvised monologues, certain segments have similar motions. To this end, to identify the segments that are the most different from the rest of the subsequences, we rely on LMA-based features. Inspired by the works of Aristidou et al. [92] and Durupinar et al. [63], we employ

distance-based LMA parameters. Our implementation of the LMA parameters is based on the joint locations using the same skeleton, employing different poses at different animation frames. Since the same skeleton is used, different poses made from a set of joint rotations can be compared regarding the distances of specific joints. For example, the distance between hands will not change due to different limb lengths, but can only change due to the motion. The parameters we use as features for our ZeroEGGS samples are as follows:

1. $D_{Hand_L Hand_R}$
2. $D_{Hand_L Head}$
3. $D_{Hand_R Head}$
4. $D_{Foot_L Foot_R}$
5. $D_{Elbow_L Elbow_R}$
6. $D_{Hips Head}$
7. $D_{Knee_L Knee_R}$
8. $D_{Hips Hand_L}$
9. $D_{Hips Hand_R}$
10. S_{Hand_L}
11. S_{Hand_R}
12. S_{Elbow_L}
13. S_{Elbow_R}
14. S_{Head}
15. S_{Neck}
16. $A_{Hand_L Elbow_L Shoulder_L}$
17. $A_{Hand_R Elbow_R Shoulder_R}$
18. $A_{Spine Chest Neck}$
19. $A_{Foot_L Knee_L Leg_L}$
20. $A_{Foot_R Knee_R Leg_R}$

These 20 features represent the personality style of each ZeroEGGS segment. These parameters are calculated for each frame to form a $N - 1 \times 20$ feature matrix for a segment with N frames due to speed features. However, since different segments vary in the number of frames, we use an approach where multiple frames are averaged to form a feature matrix of $M \times 20$, where M is a constant. For example, suppose Segment A has 60×20 feature matrix and Segment B has 30×20 ; using $M = 10$, we form a 10×20 feature matrix for each animation where six consecutive frames are averaged for each bin of Segment A, and three

consecutive frames for Segment B. Since the animations use a constant frame rate and only vary in total duration, using M bins for the features corresponds to comparable measurements.

The value of M acts as the sampling rate of the features across the time dimension. We observe that $M > 16$ increases the detail of the features but has little effect when calculating the difference between segments. In other words, for $M < 16$, we obtain useful features at different resolutions that distinguish the segments sufficiently. After some point, the higher resolutions do not contribute to more distinguishable features, and sometimes, noise interferes. The effect of using different sampling rates on the feature matrices is shown in Figure 4.3. We observe that the higher the sampling rate, the more detail the feature matrix can capture; however, for very high rates, the features also capture noise, which is not desired.

We define the difference between two segments, a and b , at a sample rate M , $\Delta_M(a, b)$, as the distance between their feature matrices at the same sample rate M . This distance is calculated element-wise, and we obtain a single number that indicates how different these two samples are. For example, $\Delta_M(a, a) = 0$ when the sample is used. The distance for certain pairs can be different for various M values; however, $\Delta_{M-i}(a, b) \approx \Delta_M(a, b)$ for $M - i > 16$. The sampling rate M cannot be larger than the number of frames of an animation.

We calculate the absolute difference between two segments, a and b , as $\Delta(a, b) = \frac{1}{16} \sum_{k=1}^{16} \Delta_k(a, b)$. In other words, we take the average of the distance for M values in the $[1, 16]$ range for the absolute difference. The sampling rate range is determined based on the ZeroEGGS segments' distance values. When the samples are similar, we observe that they result in close $\Delta(a, b)$, and vice versa. We depict an example case for the most similar pair in Figure 4.4. We only show a subset of features at a constant sampling rate to visualize the example. Notice how the two animations are very close in timing and hand motions, although they have slightly different overall postures.

We aim to eliminate similar segments so that we are left with a manageable

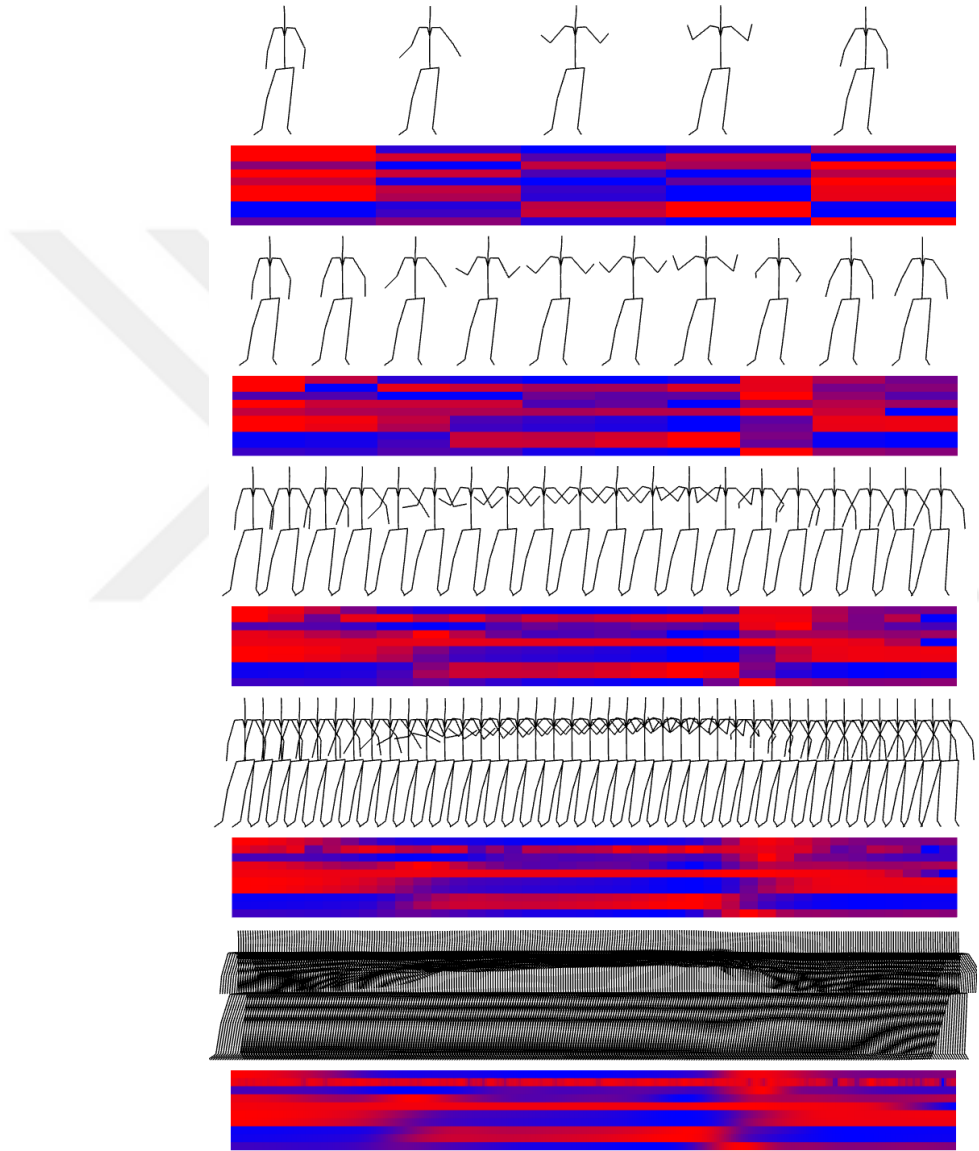


Figure 4.3: Feature vectors with different sampling rates. M values are 5, 10, 20, 40, and 80 from top to bottom. Stick figures represent the different frames of the animation segment, and the colors below the frames represent a subset of LMA feature values. Blue indicates low values, and red indicates high values. Color values are scaled based on the current range.

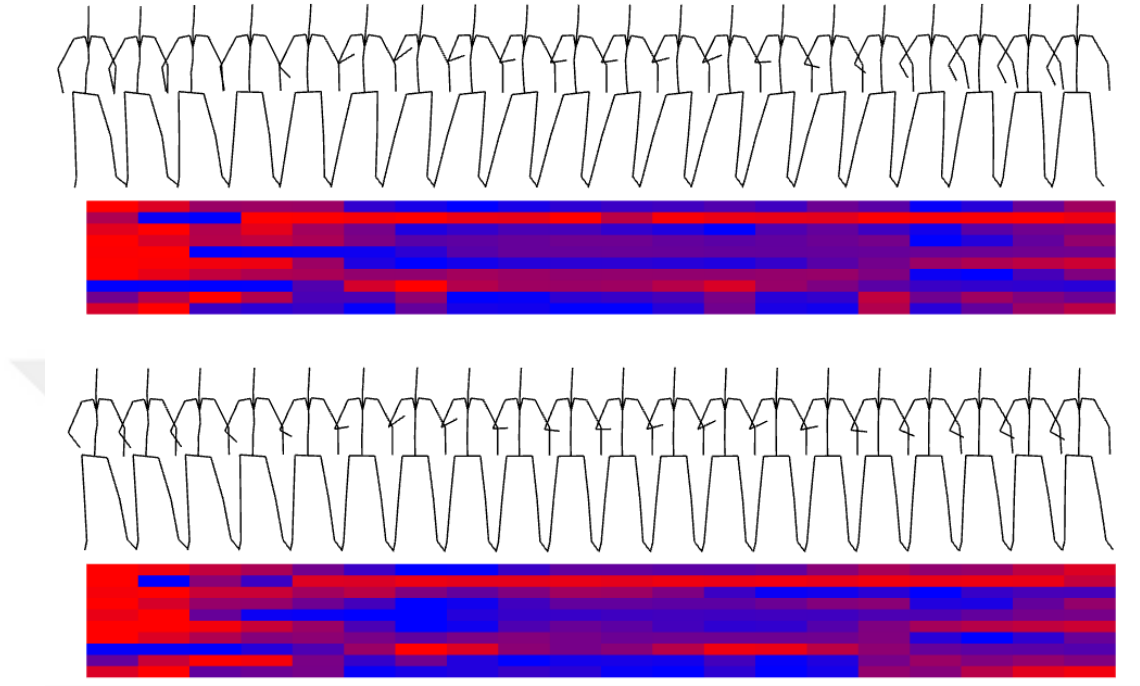


Figure 4.4: This segment pair has the smallest distance among all ZeroEGGS segments. Notice how they have similar hand movements and timing.

number of samples for the user study. For this purpose, we calculate the pairwise distance between all segments and use an iterative elimination process. Starting with all 1,291 segments, we identify the smallest distance pair. Then, we eliminate the segment with the smallest average distance from all other remaining segments; in other words, the segment is more similar to the other segments. We continue this process until we are left with 100 segments. This approach is detailed in Algorithm 1. After this elimination process, we are left with the 100 segments with the highest in-between difference based on the LMA features, which are expected to have different personality types.

Although the remaining segments have mostly unique attributes regarding their motion features, they are still expected not to cover the whole range of FFM personalities. This behavior is because the ZeroEGGS dataset is not crafted for personality expression; the variation included may not cover the whole range for all traits. Certain personality factors like openness are more difficult to perceive and express in short animation segments [64], and consequently, they may not

```

for  $segment \in segments$  do
  for  $k \in \{1, 2, \dots, 16\}$  do
     $features_k(segment) \leftarrow calculateFeatures_k(segment)$ 
  end
end
for  $i \in size(features)$  do
  for  $j \in size(features)$  do
    if  $i \neq j$  then
       $\Delta(i, j) \leftarrow \frac{1}{16} \sum_{k=1}^{16} |features_k(i) - features_k(j)|$ 
    else
       $\Delta(i, j) \leftarrow \infty$ 
    end
  end
end
while  $size(animations) > 100$  do
   $i, j = findMin(\Delta(i, j))$ 
  if  $\sum_x \Delta(i, x) > \sum_x \Delta(j, x)$ , for all remaining  $x$  samples then
     $remove(j)$ 
  else
     $remove(i)$ 
  end
end

```

Algorithm 1: The approach we use for reducing the number of ZeroEGGS segments by progressively finding and removing the most similar pairs. We remove the sample closer to the remaining segments for each such pair.

span the whole range of the related factors.

4.2.2 User Study Details

In our perception study, we display the Bandai dataset animations (128 samples) and the ZeroEGGS segments (100 samples) to the participants we acquired from the crowdsourcing platform Prolific and the email-invited subjects. We implement a Unity WebGL application to display the animations on a neutral 3D model. We do not use a realistic human model to avoid bias due to body appearance. The user interface of our annotation system is shown in Figure 4.5. The users invited by email participated voluntarily and were allowed to rate any samples. The participants from Prolific were paid based on the time spent on the study. Each Prolific participant was assigned 10 tasks that included rating a randomly selected sample. Two hundred ninety participants rated the 228 samples, where at least 15 users rated each.

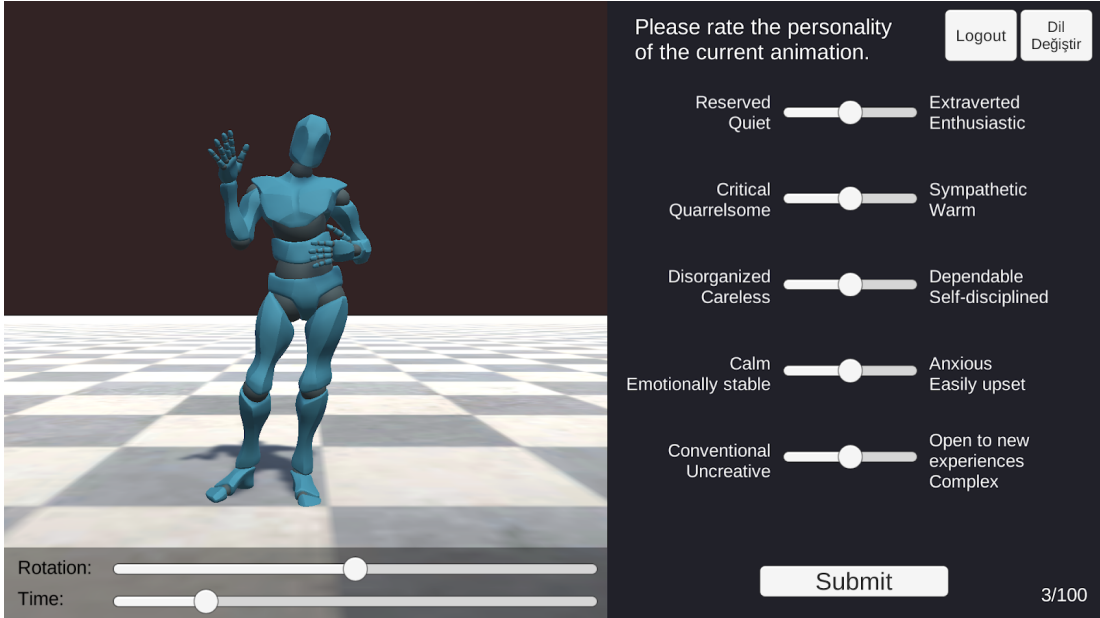


Figure 4.5: The Unity WebGL application used in annotating the Bandai and ZeroEGGS samples regarding the personality traits they express. The participants used the five sliders to rate the looping motion based on five factors.

For rating the samples, we used a short form of the Ten-Item Personality

Inventory (TIPI) [127], where each factor appears with its polar opposite on a 7-point Likert scale. Moving the sliders causes the related labels to change in size to give visual feedback to the participant. Upon clicking on the submit button, the next task is displayed. The user can rotate the camera around the figure and pause, rewind, or fast-forward the animation. The participants cannot return to previously submitted animations. We support two languages, Turkish and English, to help users better identify each trait. The participants acquired from Prolific must be fluent in English, and the participants acquired with an email invitation are native Turkish speakers. The participants' answer to each question is mapped to an integer in the $[-3, 3]$ range to be used in statistical analysis. After filtering out any indecisive annotations, we first examine the histograms for each sample and use the mean scores for the personality factors in the rest of the study.

4.3 Filtering Indecisive Annotations

Crowdsourcing perception studies require a consensus so that the results can be used as annotations to describe each sample. With such studies, we expect the histograms of the user answers to resemble a normal distribution. For example, any distribution with multiple peaks, especially if the peaks are on opposite polarities of the corresponding personality factor, cannot be used to represent the sample. On the other hand, in certain cases, we observe a partial normal distribution centered on the boundary of the measurement range, which is acceptable. For example, due to the limited measurement range, the ratings could peak at 3 or -3, which prevents observing the full range of the distribution, but shows a consensus for the measured personality factor.

To identify indecisive annotations, we first use the Shapiro-Wilk test for normality. If the resulting p-value is less than 0.05 for the corresponding personality factor, we mark that distribution for further inspection. Among these further investigated distributions, we reject any without a clear peak or with opposing sign peaks, which suggests the results have no consensus.

We use sample distributions from our Bandai annotations to illustrate the accepted and rejected distributions. Figure 4.6 depicts the subset of samples where the Shapiro-Wilk test fails for all five factors for each case. However, notice how some of these distributions fail normality only because we observe the one-tailed portion of the distribution. For example, for “Sample 7” on the first row, conscientiousness distribution (second column) has a clear consensus. However, it fails normality since we do not observe any rating larger than 3; the peak is at the boundary. We include additional conditions to accept such distributions even if the normality check fails.

Similarly, for certain cases, many of the ratings belong to the positive or the negative part, and even though the distribution does not show normality, the results depict that the sample is positive or negative. For example, in Figure 4.6, “Sample 8” fails the Shapiro-Wilk test, which is not the one-tailed half of a normal distribution. However, we observe that most ratings indicate a relatively positive perception. We describe two conditions to accept distributions even if they fail the normality check:

- **Majority Condition:** If the histogram bin with the maximum number of ratings is larger than the sum of the remaining bins, we accept these cases as valid since the selected answer has the majority. The average would be close to the majority’s answer.
- **Difference Condition:** If the sum of the histogram bins for the negative ratings and the sum of the ones for the positive ratings has an absolute difference of at least ten, we again accept such cases as valid. In this case, either positive or negative ratings have the majority, and the Shapiro-Wilk test fails since the amount of positivity or negativity is indecisive. For example, the same number of participants could have chosen minor and major agreement for the corresponding trait, and the average will still represent an acceptable value.

Any distribution that fails the normality condition but is accepted due to one of the additional conditions is shown with a green frame in Figures 4.6 and 4.7. In

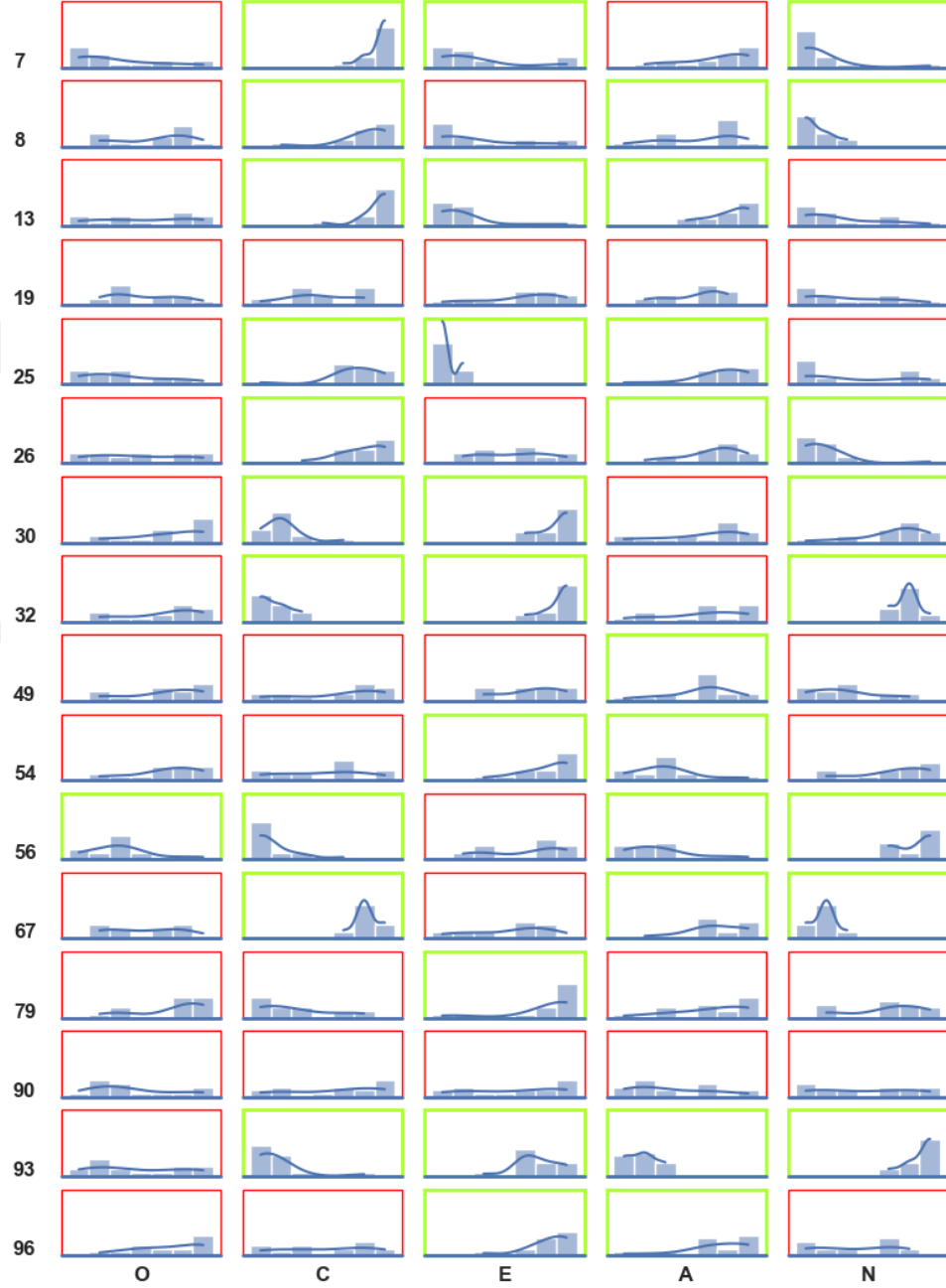


Figure 4.6: Per sample rating distributions for each personality factor for a subset of Bandai samples. This subset of samples fails the Shapiro-Wilk test for all factors. Samples with the green frame depict the accepted cases based on our additional conditions. The ones with the red frame are rejected.

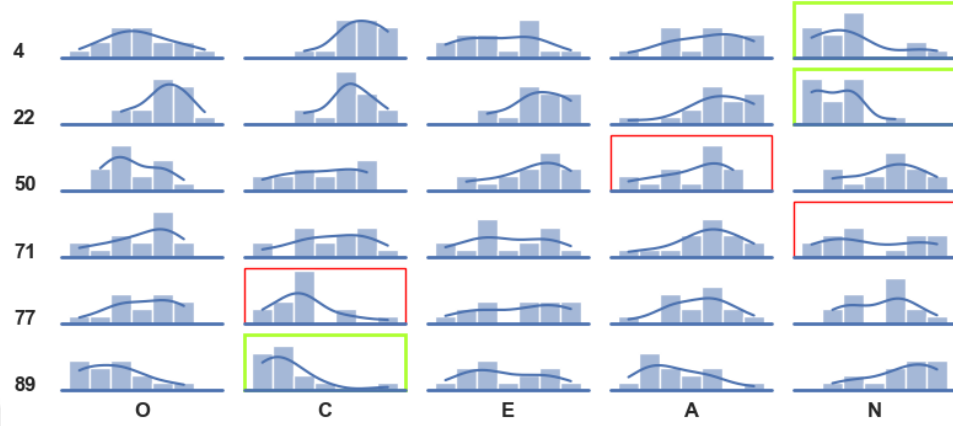


Figure 4.7: Per sample rating distributions for each personality factor for a subset of Bandai samples. This subset of samples fails the Shapiro-Wilk test for only one factor. Samples with the green frame depict the accepted cases based on our additional conditions. The ones with the red frame are rejected.

Figure 4.7, we observe the distributions where only one factor fails the normality check; these samples mostly achieve a consensus. For example, in this figure, neuroticism distributions for “Sample 4” and “Sample 22” fail the normality condition, but are accepted due to the difference condition; the difference between the negative and positive ratings for these distributions is high. On the other hand, the neuroticism distribution for “Sample 71” had a similar number of votes on the positive and negative polarity; hence, it was rejected.

We use the distribution mean for each sample as the personality annotation for the corresponding factor for any accepted case. In this case, the mean score for each factor can be in the $[-3, 3]$ range. We ignore the means if the distribution is rejected. Suppose there are any valid distributions for a sample. In that case, it is used to analyze the corresponding factor, regardless of whether it has a valid distribution for the remaining factors. For example, “Sample 79” in Figure 4.6 has an accepted distribution for extraversion, so it is used in the correlation analysis for extraversion. However, it is ignored for the rest of the personality factors as their distributions are rejected. Certain animations like “Sample 19” and “Sample 90” fail all the factors and thus are not included in the rest of the analysis.

Dataset	Case	Personality				
		O	C	E	A	N
Bandai	Accepted (Normal Distribution)	66	47	21	47	37
	Accepted (Conditions)	24	48	58	38	48
	Total Accepted	90	95	79	85	85
	Rejected	38	33	49	43	43
ZeroEGGS	Accepted (Normal Distribution)	58	70	19	53	50
	Accepted (Conditions)	17	3	47	15	13
	Total Accepted	75	73	66	68	63
	Rejected	25	27	34	32	37

Table 4.1: The number of accepted (due to having normal distribution or because of our two conditions) and rejected cases for the Bandai and ZeroEGGS dataset annotations.

Table 4.1 depicts the number of accepted and rejected distributions for each personality factor for the Bandai and ZeroEGGS annotations. Although almost a quarter of the samples are rejected in certain cases, the remaining samples are sufficient for our motion-personality analysis. Interestingly, the highest number of rejected distributions is for extraversion, the factor that is often mentioned as the easiest to perceive. This aspect, perhaps, causes mixed opinions on the perceived extraversion.

4.4 Personality Correlation Analysis

We utilize the Pearson correlation coefficients between the personality annotations and our motion features. For each motion parameter, we average the calculated parameter values over all frames of a sample to come up with a single value per sample and motion parameter. Similarly, the mean personality score represents the factor values of each sample. We analyze the correlation between these two variables, the average of the parameter and personality value, over all animation samples for each parameter-personality factor combination.

For example, suppose we have the distance between hands as our motion parameter. We calculate the average hand distance for each sample over all animation frames. Then, for each sample, we also calculate the mean extraversion score using the accepted values in the distribution. The Pearson correlation coefficient for the relation of the hand distance and extraversion corresponds to a number in the range $[-1, 1]$, where 0 refers to no correlation, 1 corresponds to a very strong correlation with direct proportion, and -1 corresponds to a very strong correlation with inverse proportion. The motion parameters that strongly correlate with personality factors are expected to have coefficients with high magnitude.

The rest of this chapter contains the best correlated motion parameters with each factor in the Bandai and ZeroEGGS datasets. We analyze the correlations for different datasets separately to favor the changes due to the actions. The Bandai dataset contains more active movements, and if we perform the correlation analysis in the union set, the more subtle differences in the ZeroEGGS samples can be overlooked. We look for the highest Pearson correlation coefficients for directly and inversely proportional parameters.

Table 4.2 shows the motion parameters that correlate well with the openness factor in the Bandai dataset. We observe that angle features dominate, especially the angles containing hand and upper body joints. Interestingly, the best parameters related to openness correlate more with extraversion. This behavior can be because short motion sequences make extraversion easier to perceive.

Table 4.3 shows the motion parameters with the best correlation coefficients for the conscientiousness factor in the Bandai dataset. Here, we observe that inversely proportional parameters include the speed of the hand and the lower arm joints. Directly proportional parameters generally involve angles. Conscientiousness parameters also highly correlate with extraversion, agreeableness, and neuroticism. Among other factors, openness seems minimally affected by the parameters influencing conscientiousness.

Table 4.4 depicts the best motion parameters related to extraversion in the Bandai dataset. While inversely proportional parameters mainly include angle

Openness – Bandai							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	$A_{Chest\ Hand_R\ LowerLeg_R}$	-.603	.106	-.680	-.264	.089
	2	$A_{Spine\ Hand_R\ LowerLeg_R}$	-.596	.060	-.647	-.290	.089
	3	$A_{LowerArm_R\ Hips\ Hand_R}$	-.591	.115	-.698	-.307	.131
	4	$A_{Neck\ Hand_R\ LowerLeg_R}$	-.572	.158	-.686	-.265	.086
	5	$A_{Spine\ Hand_R\ Toe_R}$	-.568	.088	-.642	-.255	.027
	6	$A_{Spine\ UpperArm_R\ UpperLeg_L}$	-.565	-.072	-.526	-.209	.394
	7	$A_{Chest\ LowerLeg_L\ Head}$	-.563	-.166	-.296	-.334	.402
	8	$A_{Hips\ Hand_R\ UpperLeg_L}$	-.562	.234	-.697	-.137	.129
	9	$A_{Spine\ Hand_R\ LowerLeg_L}$	-.559	.091	-.617	-.335	.136
	10	$A_{Spine\ Hand_R\ Foot_R}$	-.559	.082	-.625	-.260	.028
Directly Proportional	1	$A_{UpperArm_R\ Chest\ UpperLeg_L}$	.567	.003	.590	.217	-.296
	2	$A_{Hand_L\ Neck\ UpperArm_R}$	.561	-.293	.678	.100	.108
	3	$A_{Hand_R\ Spine\ LowerLeg_R}$	.560	-.021	.592	.320	-.081
	4	$A_{Hand_R\ Spine\ UpperLeg_L}$	.552	-.067	.606	.368	-.124
	5	$A_{Hand_R\ Spine\ Foot_R}$	.533	-.043	.578	.280	-.043
	6	$A_{Hand_R\ Spine\ LowerLeg_L}$	.526	-.046	.557	.393	-.147
	7	$A_{Hand_L\ LowerArm_L\ UpperLeg_R}$	.512	-.186	.484	.088	.082
	8	$A_{Hand_R\ Spine\ Toe_L}$	.512	-.058	.555	.366	-.115
	9	$A_{Hand_L\ Head\ Hand_R}$	.509	-.329	.739	.202	-.035
	10	$A_{Hand_L\ Neck\ Shoulder_R}$	.506	-.242	.579	.092	.135

Table 4.2: The motion parameters with the best correlation to openness in the Bandai dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

Conscientiousness – Bandai							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	S_{Hand_R}	.039	−.483	.509	−.351	.525
	2	S_{Hand_L}	.022	−.449	.451	−.397	.526
	3	$S_{LowerArm_R}$	−.006	−.422	.414	−.405	.491
	4	$S_{LowerArm_L}$	−.011	−.406	.388	−.422	.487
	5	$A_{UpperArm_L Shoulder_L Hand_R}$	.394	−.403	.583	.160	.045
	6	$A_{Chest Foot_R Toe_R}$	−.140	−.402	.262	−.499	.401
	7	$A_{Hips Foot_L LowerLeg_R}$	−.049	−.400	.404	−.514	.380
	8	$A_{Head Foot_R Toe_R}$	.030	−.397	.397	−.448	.343
	9	$A_{UpperArm_L Shoulder_R Hand_R}$	.429	−.396	.571	.197	−.002
	10	$A_{Shoulder_L Foot_R Toe_R}$	−.041	−.393	.336	−.469	.364
Directly Proportional	1	$A_{Foot_R UpperArm_R Toe_R}$	−.013	.434	−.397	.421	−.348
	2	$A_{UpperArm_R Toe_R Foot_R}$	.032	.423	−.361	.470	−.373
	3	$A_{Foot_R Shoulder_R Toe_R}$	.026	.419	−.350	.435	−.349
	4	$A_{Foot_R LowerArm_R Toe_R}$	−.157	.418	−.559	.183	−.299
	5	$A_{Shoulder_R Toe_R Foot_R}$	.045	.413	−.344	.478	−.376
	6	$A_{Foot_R Chest Toe_R}$	.154	.409	−.242	.467	−.387
	7	$A_{Foot_R Neck Toe_R}$	−.003	.408	−.363	.419	−.340
	8	$A_{Neck Toe_R Foot_R}$	.021	.404	−.359	.469	−.364
	9	$A_{Foot_R Spine Toe_R}$	.189	.398	−.217	.474	−.386
	10	$A_{Foot_R Shoulder_L Toe_R}$	.028	.397	−.331	.422	−.347

Table 4.3: The motion parameters with the best correlation to conscientiousness in the Bandai dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

Extraversion – Bandai							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	$A_{UpperArm_L Hand_R LowerLeg_R}$	-.586	.166	-.699	-.273	.092
	2	$A_{Hips Hand_R UpperLeg_L}$	-.562	.234	-.697	-.137	.129
	3	$A_{Hips LowerArm_R UpperLeg_R}$	-.504	.113	-.695	-.277	.107
	4	$A_{Hips LowerArm_R Spine}$	-.505	.121	-.688	-.289	.104
	5	$A_{Hips LowerArm_R Chest}$	-.499	.138	-.674	-.315	.099
	6	$A_{Hips LowerArm_L UpperLeg_L}$	-.503	.110	-.647	-.025	-.019
	7	$A_{Hips LowerArm_R Shoulder_L}$	-.469	.146	-.633	-.351	.100
	8	$A_{Hips LowerArm_L Spine}$	-.493	.132	-.631	-.031	-.052
	9	$A_{Hips LowerArm_L Chest}$	-.495	.147	-.630	-.065	-.048
	10	$A_{Hips LowerArm_R Shoulder_R}$	-.474	.134	-.629	-.345	.098
Directly Proportional	1	$V_{7,8,12,18,19}$	.456	-.342	.673	.021	.118
	2	$A_{UpperArm_L Neck Hand_R}$	.500	-.307	.653	.333	-.026
	3	$A_{LowerArm_R Chest LowerLeg_R}$	.503	-.132	.643	.312	-.059
	4	$V_{7,8,11,12,18,19}$	.360	-.372	.640	-.003	.228
	5	$V_{7,8,12,18}$	.379	-.229	.632	.211	.037
	6	$V_{8,11,12,18,19}$	.430	-.384	.632	-.006	.172
	7	$V_{7,8,12,14,15}$	.370	-.397	.626	-.155	.186
	8	$V_{7,8,12,14,18,19}$	.343	-.378	.620	-.261	.226
	9	$V_{7,8,11,12,14,18,19}$	.295	-.398	.613	-.273	.294
	10	$V_{7,8,11,12,14,18}$	.291	-.362	.612	-.213	.282

Table 4.4: The motion parameters with the best correlation to extraversion in the Bandai dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

measurements, volume measurements correlate highly with extraversion. The upper body joints have a dominant role in parameters that explain extraversion. We observe that the motion parameters with high magnitude correlation coefficients for extraversion similarly affect openness. This correlation is also apparent in related work, using the subgroup plasticity to combine extraversion and openness together [15].

In Table 4.5, we depict the motion parameters that highly correlate with agreeableness in the Bandai dataset. Distance-based parameters are dominant for agreeableness, followed by angle features. The distances due to the foot and toe

Agreeableness – Bandai							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	$D_{Foot_L Foot_R}$	-.001	-.307	.424	-.586	.352
	2	$D_{Toe_L Foot_R}$	-.003	-.287	.402	-.585	.355
	3	$D_{Toe_L Toe_R}$	.015	-.308	.439	-.568	.335
	4	$D_{Foot_L Toe_R}$	.026	-.314	.462	-.567	.327
	5	$D_{LowerLeg_L Toe_R}$	-.010	-.254	.426	-.550	.297
	6	$A_{Chest Toe_R LowerLeg_L}$	-.030	-.343	.387	-.548	.372
	7	$A_{Hips Toe_R LowerLeg_L}$	-.051	-.349	.397	-.542	.381
	8	$A_{Head Toe_R LowerLeg_L}$	.029	-.320	.428	-.536	.316
	9	$D_{LowerLeg_L Foot_R}$	-.028	-.238	.379	-.534	.319
	10	$A_{Chest Foot_R LowerLeg_L}$	-.055	-.358	.344	-.533	.410
Directly Proportional	1	$A_{UpperArm_L LowerLeg_L Toe_R}$	-.008	.324	-.408	.581	-.371
	2	$A_{Shoulder_L LowerLeg_L Toe_R}$	.001	.326	-.410	.574	-.365
	3	$A_{LowerArm_R LowerLeg_R Foot_L}$	-.013	.343	-.405	.566	-.363
	4	$A_{Shoulder_L LowerLeg_L Foot_R}$	.015	.339	-.377	.564	-.405
	5	$A_{LowerArm_R Foot_R Toe_L}$	.045	.255	-.268	.561	-.366
	6	$A_{LowerArm_R LowerLeg_R Toe_L}$	-.049	.318	-.424	.557	-.335
	7	$A_{Head LowerLeg_L Toe_R}$	-.008	.316	-.421	.555	-.333
	8	$A_{LowerArm_R Toe_R Toe_L}$	.016	.288	-.282	.553	-.387
	9	$A_{LowerLeg_L Foot_R Foot_L}$	.038	.276	-.336	.549	-.306
	10	$A_{UpperArm_R Toe_R Toe_L}$	.010	.329	-.390	.547	-.337

Table 4.5: The motion parameters with the best correlation to agreeableness in the Bandai dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

joints are especially essential to explain agreeableness. The parameters correlating with agreeableness also affect perceived conscientiousness, extraversion, and neuroticism. The relationship between agreeableness and extraversion is inverse. Low agreeableness translates into high neuroticism and extraversion; agreeable individuals tend to have less dramatic movements.

Table 4.6 shows the motion parameters that correlate well with neuroticism in the Bandai dataset. We observe that acceleration-based measurements best relate to neuroticisms, followed by angle and distance-related features. The inverse proportional relationship between neuroticism and agreeableness is also apparent in these results. Openness is rather isolated, and the parameters that explain

Neuroticism – Bandai							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	$A_{Hips\ LowerLeg_R\ Toe_R}$	.113	.333	-.263	.512	-.448
	2	$A_{Hips\ LowerLeg_L\ Foot_L}$	.158	.347	-.187	.501	-.447
	3	$A_{UpperLeg_L\ LowerLeg_L\ Toe_L}$	.132	.341	-.195	.480	-.444
	4	$A_{Chest\ LowerLeg_R\ Foot_R}$	.085	.300	-.217	.506	-.435
	5	$D_{UpperLeg_R\ Foot_R}$	.097	.305	-.252	.460	-.435
	6	$D_{UpperLeg_L\ Foot_L}$	.102	.349	-.234	.445	-.434
	7	$A_{Spine\ LowerLeg_L\ Foot_L}$	.128	.352	-.196	.502	-.430
	8	$A_{Neck\ Chest\ LowerLeg_R}$	.539	.174	.278	.365	-.426
	9	$A_{Chest\ LowerLeg_R\ Toe_R}$	.054	.312	-.267	.508	-.424
	10	$A_{Chest\ LowerLeg_L\ Foot_R}$	.060	.346	-.356	.562	-.420
Directly Proportional	1	C_{Hand_L}	-.013	-.529	.465	-.360	.580
	2	$C_{UpperLeg_R}$	-.078	-.472	.401	-.598	.565
	3	$C_{UpperLeg_L}$	-.076	-.481	.406	-.590	.563
	4	C_{Hips}	-.071	-.471	.402	-.600	.557
	5	$C_{LowerArm_R}$	.014	-.485	.491	-.333	.549
	6	$C_{LowerArm_L}$	-.011	-.457	.437	-.383	.541
	7	C_{Head}	-.050	-.435	.434	-.505	.532
	8	C_{Hand_R}	.048	-.518	.548	-.245	.528
	9	$C_{Shoulder_R}$	-.023	-.418	.433	-.511	.524
	10	$C_{UpperArm_L}$	-.011	-.418	.433	-.496	.523

Table 4.6: The motion parameters with the best correlation to neuroticism in the Bandai dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

neuroticism have high correlations with the remaining factors.

Table 4.7 shows the motion parameters with the best openness correlations in the ZeroEGGS dataset. Unlike in the Bandai dataset, this time, the parameters that explain openness also relate to agreeableness, neuroticism, and extraversion. We observe that angle-based parameters are mostly effective, with upper body joints more influential. The parameters with inverse proportion have higher magnitude correlation coefficients with extraversion than openness, similar to the Bandai dataset. Openness and extraversion have a direct proportion.

Openness – ZeroEGGS							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	$A_{Chest\ Hips\ Head}$	−.667	.110	−.713	−.570	.469
	2	$A_{Hips\ Head\ Chest}$	−.665	.113	−.713	−.565	.457
	3	$A_{Chest\ UpperLeg_R\ Head}$	−.665	.085	−.640	−.431	.390
	4	$A_{Hips\ Head\ UpperLeg_R}$	−.664	.091	−.687	−.376	.428
	5	$A_{Hips\ Head\ Spine}$	−.662	.065	−.695	−.551	.519
	6	$A_{Shoulder_L\ UpperLeg_R\ UpperArm_R}$	−.656	.190	−.633	−.455	.403
	7	$A_{Chest\ Head\ UpperLeg_R}$	−.653	.075	−.648	−.425	.376
	8	$A_{Spine\ Head\ UpperLeg_R}$	−.651	.100	−.579	−.332	.380
	9	$A_{Spine\ Hips\ Head}$	−.649	.057	−.681	−.541	.525
	10	$A_{Chest\ UpperLeg_R\ Shoulder_L}$	−.640	.109	−.682	−.482	.428
Directly Proportional	1	$A_{Hips\ Chest\ Head}$	.666	−.111	.713	.568	−.464
	2	$A_{Head\ Chest\ UpperLeg_R}$	.660	−.081	.644	.429	−.384
	3	$A_{Shoulder_L\ Chest\ UpperLeg_R}$	.636	−.104	.683	.487	−.426
	4	$A_{UpperArm_R\ Spine\ UpperLeg_R}$	.630	−.149	.604	.358	−.389
	5	$A_{Head\ Spine\ UpperLeg_R}$	.629	−.096	.499	.305	−.353
	6	$A_{Shoulder_L\ Spine\ UpperLeg_R}$	.628	−.145	.592	.340	−.389
	7	$A_{Neck\ Spine\ UpperLeg_R}$	.627	−.135	.595	.338	−.383
	8	$A_{Shoulder_R\ Spine\ UpperLeg_R}$	.625	−.150	.593	.334	−.385
	9	$A_{Head\ Shoulder_L\ Toe_L}$	.612	−.039	.617	.391	−.426
	10	$A_{Shoulder_R\ Chest\ UpperLeg_R}$	.612	−.101	.649	.407	−.370

Table 4.7: The motion parameters with the best correlation to openness in the ZeroEGGS dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

Conscientiousness – ZeroEGGS							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	S_{Hand_L}	.514	−.395	.677	.301	−.175
	2	S_{Hand_R}	.484	−.375	.766	.183	−.111
	3	$S_{LowerArm_R}$	.500	−.362	.776	.147	−.089
	4	$S_{LowerArm_L}$	.576	−.350	.744	.255	−.140
	5	$D_{Foot_L Toe_L}$	.287	−.305	.224	.174	−.013
	6	S_{Head}	.395	−.300	.535	.246	−.074
	7	$S_{UpperArm_R}$	.493	−.291	.631	.280	−.105
	8	$A_{Shoulder_L Toe_R LowerLeg_R}$	.383	−.289	.303	.235	−.142
	9	S_{Neck}	.424	−.283	.566	.246	−.066
	10	$A_{Neck Toe_R LowerLeg_R}$	.460	−.276	.382	.308	−.221
Directly Proportional	1	$D_{Hips Spine}$	−.253	.324	−.241	−.008	.048
	2	$A_{Toe_L Foot_R LowerLeg_R}$	−.035	.292	−.066	.087	−.079
	3	$A_{Shoulder_L LowerLeg_R Toe_R}$	−.362	.285	−.295	−.217	.122
	4	$A_{Neck LowerLeg_R Toe_R}$	−.442	.273	−.374	−.294	.204
	5	$A_{UpperArm_L LowerLeg_R Toe_R}$	−.200	.262	−.198	−.139	.003
	6	$A_{Head Hand_L UpperArm_R}$	−.228	.262	−.360	−.231	.157
	7	$A_{Shoulder_R LowerLeg_R Toe_R}$	−.409	.259	−.329	−.236	.159
	8	$A_{UpperArm_R Head LowerArm_R}$	−.207	.254	−.196	.011	.082
	9	$A_{Head LowerArm_L UpperArm_L}$	.046	.250	.044	.004	−.049
	10	$A_{Foot_L Foot_R LowerLeg_R}$	−.228	.249	−.207	−.053	.112

Table 4.8: The motion parameters with the best correlation to conscientiousness in the ZeroEGGS dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

Table 4.8 shows the motion parameters that effectively relate to conscientiousness in the ZeroEGGS dataset. Like the Bandai dataset, speed-related parameters have a dominant effect in addition to the angle-based parameters. However, in general, the magnitudes of the correlation coefficients are very small in contrast to the parameters of the Bandai dataset. We observe that certain parameters have moderate to strong correlation to extraversion, but the same parameters have very weak correlation to conscientiousness.

Table 4.9 depicts the strong correlations between our motion parameters and extraversion in the ZeroEGGS dataset. Angle-based parameters excel at inversely

Extraversion – ZeroEGGS							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	$A_{Hips\ Head\ Chest}$	-.665	.113	-.713	-.565	.457
	2	$A_{Chest\ Hips\ Head}$	-.667	.110	-.713	-.570	.469
	3	$A_{Hips\ Head\ Spine}$	-.662	.065	-.695	-.551	.519
	4	$A_{Head\ LowerLeg_L\ Shoulder_R}$	-.565	.064	-.693	-.449	.451
	5	$A_{Hips\ Shoulder_R\ Chest}$	-.634	.101	-.692	-.548	.476
	6	$A_{Chest\ Hips\ Shoulder_R}$	-.632	.101	-.691	-.544	.482
	7	$A_{Chest\ Spine\ Head}$	-.636	.141	-.691	-.546	.423
	8	$A_{Spine\ Head\ Chest}$	-.633	.141	-.691	-.539	.417
	9	$A_{Neck\ LowerLeg_L\ Head}$	-.584	.070	-.688	-.403	.400
	10	$A_{Chest\ Spine\ Shoulder_R}$	-.619	.127	-.687	-.542	.429
Directly Proportional	1	$C_{LowerArm_R}$	.356	-.341	.695	-.044	.014
	2	$C_{LowerArm_L}$	.506	-.301	.690	.211	-.091
	3	$A_{Shoulder_L\ Chest\ UpperLeg_R}$	.636	-.104	.683	.487	-.426
	4	$C_{UpperArm_R}$	.553	-.205	.658	.257	-.125
	5	$C_{UpperArm_L}$	.560	-.195	.654	.263	-.129
	6	$A_{Shoulder_R\ Chest\ UpperLeg_R}$	.612	-.101	.649	.407	-.370
	7	$C_{UpperLeg_R}$	.539	-.209	.637	.213	-.095
	8	C_{Hand_R}	.306	-.356	.636	-.078	.021
	9	$C_{UpperLeg_L}$	.575	-.194	.635	.283	-.136
	10	$A_{UpperArm_R\ Chest\ UpperLeg_R}$	.586	-.088	.614	.401	-.334

Table 4.9: The motion parameters with the best correlation to extraversion in the ZeroEGGS dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

proportional correlations, while acceleration measurements dominate directly proportional relationships. These parameters also highly correlate with openness and have a weak relationship to agreeableness and neuroticism. As the most common factor, extraversion achieves our analysis’s highest correlation coefficient magnitudes.

Table 4.10 includes the highest-ranked motion parameters that explain agreeableness in the ZeroEGGS dataset. We observe that angle-based features determine agreeableness. Unlike in the Bandai dataset, extraversion and agreeableness are directly proportional in the ZeroEGGS dataset. The features that explain

Agreeableness – ZeroEGGS							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	$A_{Hips\ UpperArm_L\ UpperLeg_L}$	-.513	.167	-.495	-.627	.547
	2	$A_{Shoulder_L\ UpperLeg_L\ UpperArm_L}$	-.470	.086	-.431	-.613	.532
	3	$A_{Hips\ UpperArm_L\ Chest}$	-.481	.098	-.463	-.611	.562
	4	$A_{Chest\ UpperArm_L\ UpperLeg_R}$	-.558	.071	-.572	-.596	.543
	5	$A_{Shoulder_L\ Hips\ UpperArm_L}$	-.478	.187	-.448	-.596	.488
	6	$A_{Chest\ UpperArm_L\ UpperLeg_L}$	-.399	.110	-.376	-.585	.527
	7	$A_{UpperArm_L\ Hips\ Shoulder_R}$	-.500	.187	-.487	-.580	.481
	8	$A_{Chest\ UpperLeg_L\ Head}$	-.538	.088	-.616	-.579	.469
	9	$A_{Spine\ UpperArm_L\ Chest}$	-.415	.146	-.383	-.577	.514
	10	$A_{Hips\ Shoulder_L\ Chest}$	-.594	.100	-.631	-.573	.520
Directly Proportional	1	$D_{Hips\ UpperArm_L}$	.519	-.187	.501	.634	-.540
	2	$A_{Head\ Chest\ UpperLeg_L}$	.544	-.095	.621	.577	-.463
	3	$A_{UpperArm_L\ Chest\ UpperLeg_R}$	.553	-.065	.569	.574	-.538
	4	$A_{UpperArm_L\ Chest\ UpperLeg_L}$	.383	-.061	.358	.556	-.536
	5	$A_{Shoulder_R\ Chest\ UpperLeg_L}$	.495	-.071	.528	.545	-.528
	6	$A_{UpperArm_L\ Shoulder_L\ LowerLeg_L}$	.459	-.190	.425	.544	-.502
	7	$A_{UpperArm_L\ Shoulder_R\ LowerLeg_L}$	.456	-.182	.413	.528	-.512
	8	$A_{UpperArm_R\ Chest\ UpperLeg_L}$	.545	-.033	.613	.518	-.487
	9	$A_{LowerArm_L\ Spine\ UpperLeg_L}$	.303	.089	.434	.504	-.506
	10	$A_{Head\ Shoulder_R\ Toe_L}$	.593	-.062	.670	.456	-.443

Table 4.10: The motion parameters with the best correlation to agreeableness in the ZeroEGGS dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

agreeableness also inversely relate to openness and extraversion and have a direct proportionality with neuroticism. The lower body joints have little effect on perceived agreeableness in this dataset.

Table 4.11 depicts the motion parameters with the highest magnitude correlation coefficients with the neuroticism annotations in the ZeroEGGS dataset. Angle and distance-based parameters are dominant for this factor. The parameters with inverse proportional coefficients focus mainly on the upper body joints, while the lower body joints appear as part of directly proportional measurements. We observe that neuroticism has an inverse relationship to the remaining

Neuroticism – ZeroEGGS							
Type	Rank	Motion Parameter	Pearson Correlation				
			O	C	E	A	N
Inversely Proportional	1	$A_{UpperArm_L Chest Toe_L}$	.507	.002	.439	.399	-.571
	2	$A_{Hips Chest UpperArm_L}$	.476	-.076	.464	.584	-.564
	3	$A_{LowerArm_L Spine Toe_L}$	.464	.054	.505	.372	-.549
	4	$D_{Hips Shoulder_L}$	.556	-.085	.605	.505	-.545
	5	$A_{Neck Shoulder_R Toe_L}$	.464	.029	.392	.355	-.544
	6	$D_{Hips UpperArm_L}$	.519	-.187	.501	.634	-.540
	7	$A_{Hips Spine UpperArm_L}$	.491	.010	.517	.492	-.535
	8	$D_{LowerArm_L Toe_L}$	.403	.084	.456	.399	-.535
	9	$D_{LowerArm_L LowerLeg_L}$	.390	.062	.465	.386	-.534
	10	$D_{UpperArm_L UpperLeg_L}$	.403	-.119	.392	.618	-.531
Directly Proportional	1	$A_{Chest UpperArm_L Toe_L}$	-.515	.007	-.444	-.412	.575
	2	$A_{Shoulder_R Neck Toe_L}$	-.464	-.028	-.391	-.355	.544
	3	$A_{Shoulder_L UpperLeg_L UpperArm_L}$	-.470	.086	-.431	-.613	.532
	4	$A_{Shoulder_R Neck LowerLeg_L}$	-.462	.020	-.439	-.354	.531
	5	$A_{LowerLeg_L LowerArm_L UpperLeg_R}$	-.482	.005	-.548	-.362	.527
	6	$A_{Shoulder_R Neck Foot_L}$	-.406	-.020	-.328	-.315	.519
	7	$A_{Shoulder_L Hips Shoulder_R}$	-.556	.053	-.617	-.458	.519
	8	$A_{Toe_L LowerArm_L UpperLeg_R}$	-.455	-.044	-.502	-.354	.517
	9	$A_{Shoulder_L UpperLeg_R Shoulder_R}$	-.602	.158	-.600	-.456	.513
	10	$A_{UpperArm_L UpperLeg_L Shoulder_R}$	-.486	.078	-.482	-.594	.507

Table 4.11: The motion parameters with the best correlation to neuroticism in the ZeroEGGS dataset samples. We report the best ten inversely proportional (first ten rows) and the best ten directly proportional (second ten rows) parameters. The correlation coefficients for the other factors are also included for comparison.

factors, a known effect due to the high polarity of neuroticism having negative connotations. Certain parameters of neuroticism have higher correlations with extraversion, agreeableness, and occasionally openness.

Using the correlation results, we select a set of motion parameters for our neural model. Both for the Bandai and ZeroEGGS datasets, we select the five parameters with the highest correlation coefficient magnitude for each personality factor. The selected parameters and their correlation coefficients for each personality factor and dataset are shown in Table 4.12. If a specific motion parameter is already selected for a personality factor and dataset combination, it is not repeated in the table. In such cases, the already selected parameters are skipped for the remaining

factor-dataset combinations, and fewer parameters are included. Consequently, rather than $5 \times 5 \times 2 = 50$ parameters (five factors, five parameters for each factor, and two datasets), we have 40 motion parameters in total.

The best parameters vary among the different datasets for certain factors. For example, the best extraversion parameters for the Bandai dataset are mostly focused on the hands and have a very weak correlation to extraversion in the ZeroEGGS dataset. The best extraversion parameters selected for the ZeroEGGS dataset, on the other hand, mainly focus on the spine and neck angles. The correlations between the factors also change among the datasets. The parameters selected for the ZeroEGGS dataset show higher correlation between the perceived factors.

Parts of our motion parameters study were published as *Towards understanding personality expression via body motion* [28] in *IEEE Conference on Virtual Reality* in 2024 and as *Human Movement Personality Detection Parameters* [30] in *Signal Processing and Communications Applications Conference* in 2024. In these publications, we do not exclude personality annotations based on the normality of the distributions and analyze the parameters using all the participant data. We then show that utilizing these motion parameters benefits personality estimation in contrast to using raw pose data [30]. We utilize this pose-based personality estimation network to generate predicted personality annotations for our unlabeled samples.

Motion Parameter	Bandai					ZeroEGGS				
	O	C	E	A	N	O	C	E	A	N
$A_{Chest\ Hand_R\ LowerLeg_R}$	-.603	.106	-.680	-.264	.089	-.252	-.058	-.397	-.256	.372
$A_{Spine\ Hand_R\ LowerLeg_R}$	-.596	.060	-.647	-.290	.089	-.286	-.057	-.408	-.282	.401
$A_{LowerArm_R\ Hips\ Hand_R}$	-.591	.115	-.698	-.307	.131	-.262	-.081	-.398	-.222	.331
$A_{UpperArm_L\ Hand_R\ LowerLeg_R}$	-.586	.166	-.699	-.273	.092	-.283	-.075	-.427	-.266	.395
$A_{LowerArm_L\ UpperLeg_R\ Hand_L}$	-.586	.121	-.656	-.087	-.001	-.248	-.164	-.268	-.279	.357
C_{Hand_L}	-.013	-.529	.465	-.360	.580	.430	-.366	.598	.152	-.052
C_{Hand_R}	.048	-.518	.548	-.245	.528	.306	-.356	.636	-.078	.021
$C_{LowerArm_R}$	.014	-.485	.491	-.333	.549	.356	-.341	.695	-.044	.014
S_{Hand_R}	.039	-.483	.509	-.351	.525	.484	-.375	.766	.183	-.111
$C_{UpperLeg_L}$	-.076	-.481	.406	-.590	.563	.575	-.194	.635	.283	-.136
$A_{Hand_L\ Head\ LowerArm_R}$	.501	-.318	.774	.201	.062	.290	.029	.358	.269	-.382
$A_{Hand_L\ Neck\ LowerArm_R}$	.459	-.318	.770	.131	.063	.311	.054	.402	.274	-.379
$A_{Hand_L\ UpperArm_L\ LowerArm_R}$	.463	-.332	.766	.079	.034	.299	.016	.402	.251	-.359
$A_{Hand_L\ Shoulder_L\ LowerArm_R}$	.442	-.324	.760	.083	.052	.293	.069	.402	.264	-.372
$A_{Hand_L\ Head\ Hand_R}$	.509	-.329	.739	.202	-.035	.205	.027	.352	.233	-.346
$A_{Toe_L\ Hand_R\ Foot_R}$	-.196	-.251	.145	-.672	.378	-.300	-.162	-.304	-.306	.424
$A_{Foot_L\ Hand_R\ Foot_R}$	-.179	-.281	.191	-.667	.372	-.254	-.136	-.277	-.254	.382
$A_{Toe_L\ LowerArm_R\ Foot_R}$	-.116	-.257	.259	-.665	.382	-.246	-.164	-.210	-.264	.325
$A_{Foot_L\ LowerArm_R\ Foot_R}$	-.108	-.279	.295	-.661	.372	-.200	-.134	-.188	-.223	.298
$A_{LowerLeg_L\ LowerArm_R\ Foot_R}$	-.285	-.272	.124	-.660	.378	-.323	-.078	-.356	-.314	.390
$C_{UpperLeg_R}$	-.078	-.472	.401	-.598	.565	.539	-.209	.637	.213	-.095
C_{Hips}	-.071	-.471	.402	-.600	.557	.555	-.175	.552	.350	-.140
$A_{Chest\ Hips\ Head}$	-.521	-.162	-.409	-.158	.377	-.667	.110	-.713	-.570	.469
$A_{Hips\ Chest\ Head}$	.521	.164	.408	.158	-.377	.666	-.111	.713	.568	-.464
$A_{Hips\ Head\ Chest}$	-.521	-.165	-.407	-.158	.378	-.665	.113	-.713	-.565	.457
$A_{Chest\ UpperLeg_R\ Head}$	-.502	-.114	-.476	-.123	.352	-.665	.085	-.640	-.431	.390
$A_{Hips\ Head\ UpperLeg_R}$	-.372	-.078	-.404	-.077	.271	-.664	.091	-.687	-.376	.428
S_{Hand_L}	.022	-.449	.451	-.397	.526	.514	-.395	.677	.301	-.175
$S_{LowerArm_R}$	-.006	-.422	.414	-.405	.491	.500	-.362	.776	.147	-.089
$S_{LowerArm_L}$	-.011	-.406	.388	-.422	.487	.576	-.350	.744	.255	-.140
$D_{Hips\ Head}$	.464	.175	.401	.122	-.344	.639	-.112	.713	.471	-.454
$D_{Hips\ UpperArm_L}$	.359	-.032	.300	.049	.098	.519	-.187	.501	.634	-.540
$A_{Hips\ UpperArm_L\ UpperLeg_L}$	-.322	.026	-.263	-.007	-.104	-.513	.167	-.495	-.627	.547
$D_{UpperArm_L\ UpperLeg_L}$	.332	-.025	.266	.020	.103	.403	-.119	.392	.618	-.531
$A_{Shoulder_L\ UpperLeg_L\ UpperArm_L}$	-.353	.052	-.303	-.087	-.080	-.470	.086	-.431	-.613	.532
$A_{Hips\ UpperArm_L\ Chest}$	-.381	-.032	-.293	-.022	.029	-.481	.098	-.463	-.611	.562
$A_{Chest\ UpperArm_L\ Toe_L}$	-.293	.045	-.184	-.011	-.071	-.515	.007	-.444	-.412	.575
$A_{UpperArm_L\ Chest\ Toe_L}$	.297	-.034	.175	.025	.054	.507	.002	.439	.399	-.571
$A_{Hips\ Chest\ UpperArm_L}$	.380	.045	.289	.019	-.060	.476	-.076	.464	.584	-.564
$A_{Hips\ UpperArm_L\ Spine}$	-.385	-.072	-.290	-.045	.121	-.523	.011	-.546	-.549	.560

Table 4.12: The selected motion parameters with high correlation to at least one factor and their correlation coefficients for each personality factor and dataset.

Chapter 5

Target-Sensitive Interpersonal Animation Generation

This chapter explains our target-sensitive animation generation model pipeline that generates the main agent’s motion given the current pose of the secondary agent and the previous poses of both agents. We first summarize the aim of the model and its training process. Then we introduce the notation used in the model and the design of its architecture. This chapter also includes snapshots from sample outputs and our user study-based evaluation.

5.1 The Aim

To generate target-sensitive animation, we utilize a data-driven solution focusing on individual poses. Specifically, we use the Transformer architecture to synthesize the current pose of the main agent given the previous and current poses of the secondary agent, the previous pose of the main agent, and the target personality factors. The generated output pose for the main agent complements the current pose of the secondary agent while expressing the target personality factors.

We utilize the Inter-X [21] dataset for training our model with dyadic animations. The Inter-X dataset includes ground truth personality annotations of the performing subjects. The annotations use the FFM personality dimensions and correspond to a floating number in each factor’s [0,1] range. The dataset is captured using 89 unique subjects in different dyadic scenarios. The same subject can appear in multiple animations with different action categories or matching subjects. For example, Subject A can have three different variations for the handshake motion with Subject B, two for the handshake motion with Subject C, and four for the high-five motion with Subject B. However, since the personality annotations use self-reported questionnaire results, the personality for Subject A will be the same in all these samples, which limits the personality variance in this dataset. Consequently, we include samples from Bandai [22] and ZeroEGGS [23] datasets to introduce more personality variance during training. However, the animation samples of these two datasets include single-agent actions and, therefore, cannot be used directly in the dyadic animation synthesis pipeline; we use these samples for learning the personality aspect of motion only.

We summarize the system’s high-level working logic in Figure 5.1. The data-driven motion generation system functions on consecutive animation poses, inputting the secondary and main agent’s previous frame poses and the secondary agent’s current pose. The system also inputs the target personality in terms of five orthogonal factors. The system’s output is the main agent’s current pose, expressing the desired personality traits while complementing the motion of the secondary agent. We also require the output pose to be human-like and consistent regarding the consecutive frames of the animation. Given that the secondary agent’s motion does not include sudden jumps, inputting the consecutive frames of the secondary agent’s motion will output poses for the main agent that can be combined into a smooth animation.

Note that we input the previous frame pose of the main agent to generate the current frame’s pose; this is required to produce a smooth animation that does not change drastically in each frame; however, it also necessitates an initial pose for the main agent’s motion. During inference, we use the ground truth pose to generate the first current pose of the main agent. Then, each consecutive frame

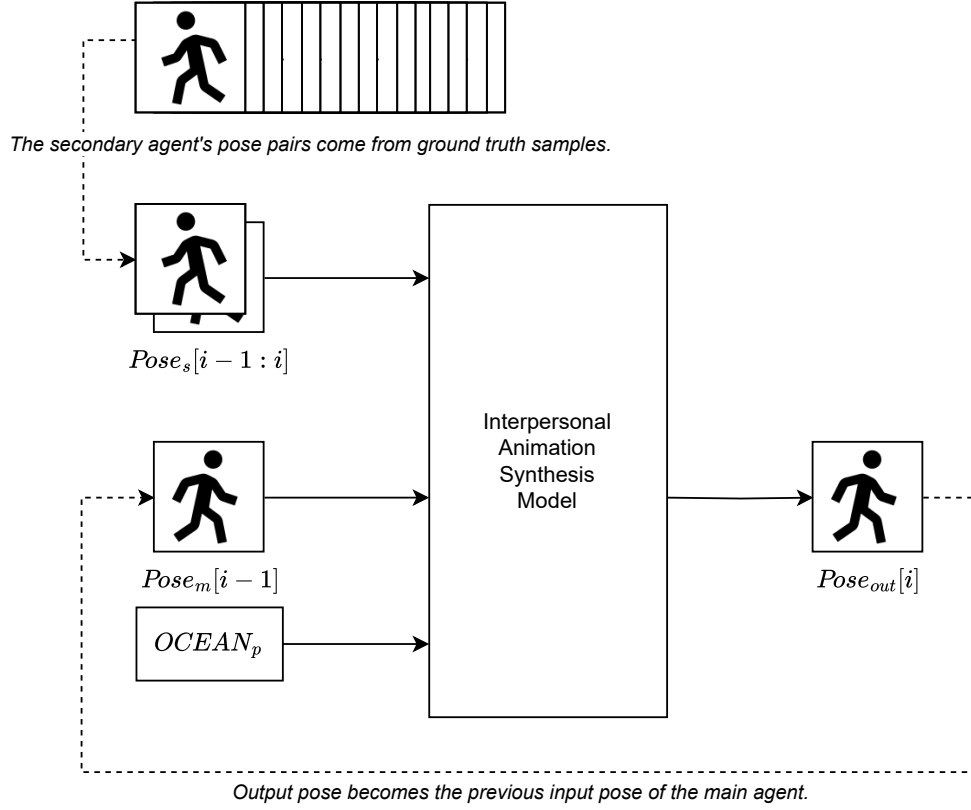


Figure 5.1: A high-level summary of our interpersonal motion synthesis model. The neural model inputs the current and previous poses for the secondary agent, the previous pose of the main agent, and the desired personality parameters to output the current pose for the main agent.

uses the output pose from the previous frame for the previous pose of the main agent. The inputs regarding the secondary agent always use the ground truth frames so that a consistent stream drives the output. This way, even if certain outputs have lower consistency, this would not accumulate while generating further frames of the same action, and the network can utilize the secondary agent’s motion to synthesize consistent poses for the main agent. The secondary agent’s poses drive the output while the previous pose of the main agent is used to achieve continuity between the main agent’s poses.

We use FFM personality factors to input the desired traits. Each factor is a floating number in the $[0, 1]$ range, where zero corresponds to the low polarity of the factor and one corresponds to the high polarity. All the samples we use in training from the Inter-X, Bandai, and ZeroEGGS datasets have personality annotations using the same format. However, since only the Inter-X dataset includes dyadic animations, we utilize the samples from the Bandai and ZeroEGGS datasets differently. During training, each batch of samples includes the network’s ground truth input and output features (previous and current pose of the secondary agent and the previous pose of the main agent to be used as input, and the ground truth current pose of the main agent to be used to calculate loss), and additionally, a randomly selected pose pair from any of the three datasets to be used for the personality loss. We use the motion features that influence personality perception to calculate the personality features using the main agent’s poses and the randomly selected pose pair. During training, we use the randomly selected sample’s personality as input to the network and expect the model to generate an output that resembles the pose of this personality sample in terms of its personality motion features.

Overall, dyadic samples are always paired and used as input to the network. However, the single-agent samples are used only for their personality values, and the corresponding poses of the single-agent samples are used in calculating the personality loss. Consequently, the network learns to output poses that satisfy the target personality in terms of previously determined motion features, without trying to replicate those poses directly. This behavior helps maintain the semantic

meaning of the motion, preventing different personality input values from dramatically changing the output. The personality input only affects certain stylistic personality choices within the range of motion variation.

5.2 The Notation

$Pose_x$ is a vector of 21×3 that represents a pose of the agent using the global positions of 21 joints. We replace x with m for the main agent, s for the secondary agent, and p for the personality agent. For example, $Pose_m$ represents one pose of the main agent corresponding to a single animation frame from the corresponding dataset we use in training. Since only the Inter-X samples contain dyadic motion, any motion pair $\{Pose_m, Pose_s\}$ is selected from Inter-X. We select $Pose_p$ samples among all three datasets since dyadic samples can also be used as two single-agent animations.

An animation sequence of N frames contains N individual poses. A specific pose that comes from a particular animation frame i is shown as $Pose_x[i]$, for $i \in \{0, 1, \dots, N-1, N\}$. We use $Pose_x[i]$ for the current pose and $Pose_x[i-1]$ for the previous pose of the agent. Our current model considers the current pose and the previous pose during training and inference; however, the architecture can be expanded to support inputting the last y number of frames, in which case $Pose_x[i-y]$ can represent one of the earlier poses.

Each sample has a corresponding personality vector of size 5, shown as $OCEAN_x$. During training, the personality of the dyadic sample $\{Pose_m, Pose_s\}$, namely $\{OCEAN_m, OCEAN_s\}$, is not fed to the network directly. Instead, we input $OCEAN_p$, the personality the network aims to express. $OCEAN_x$ of a sample does not change based on the frame, so for a specific sample's $Pose_x[i]$ and $Pose_x[i-1]$ corresponds to the same $OCEAN_x$, and therefore we do not use a time index for the personality values.

We use motion features to model the personality and interpersonal constraints

of animations. $Personality_x[i]$ symbolizes the personality features of a sample for frame i , which contains a set of distances and angles based on the current configuration of the joint positions. For example, we aim the network’s output $Pose_{out}[i]$ to be similar to $Pose_p[i]$ in terms of its personality; therefore, $Personality_{out}[i]$ should be similar to $Personality_p[i]$ during training. The interpersonal constraints are a set of distance features calculated over $\{Pose_m[i], Pose_s[i]\}$ and are symbolized as $Constraints_{ms}[i]$. These constraints should be satisfied by the network’s output so that the generated motion complements the secondary agent’s motion. For example, the hand of the main agent should be at a specific distance from the secondary agent’s hand so that the resulting animation reads as a handshake.

During training, each sample in a batch contains a paired $\{Pose_m[i - 1 : i], Pose_s[i - 1 : i]\}$ from Inter-X, that is the two consecutive frames of both agents and an unpaired $Pose_p[i - 1 : i]$ to be used for learning the personality aspect. Each sample also contains a personality value $OCEAN_x$. We feed $Pose_m[i - 1]$, $Pose_s[i - 1 : i]$, and $OCEAN_p$ to the network to receive $Pose_{out}[i]$. The output $Pose_{out}[i]$ should be similar to $Pose_m[i]$ in terms of the interpersonal constraints and pose while expressing the traits of $OCEAN_p$.

5.3 The Losses

We use a combination of losses (see Figure 5.2) during training to support the aforementioned functionality in an end-to-end solution:

- **Pose Loss ($\mathcal{L}_{pose}$):** This loss assures the generated output for the main agent’s current pose, $Pose_{out}[i]$, is similar to the ground truth pose of the main agent for the current frame $Pose_m[i]$, without considering any personality features. We use Mean Squared Error (MSE) to formulate this loss between $Pose_{out}[i]$ and $Pose_m[i]$. Each distance in the feature vector is calculated individually instead of using the Euclidean distance. Training the network using this loss only is sufficient to generate the main agent’s poses,

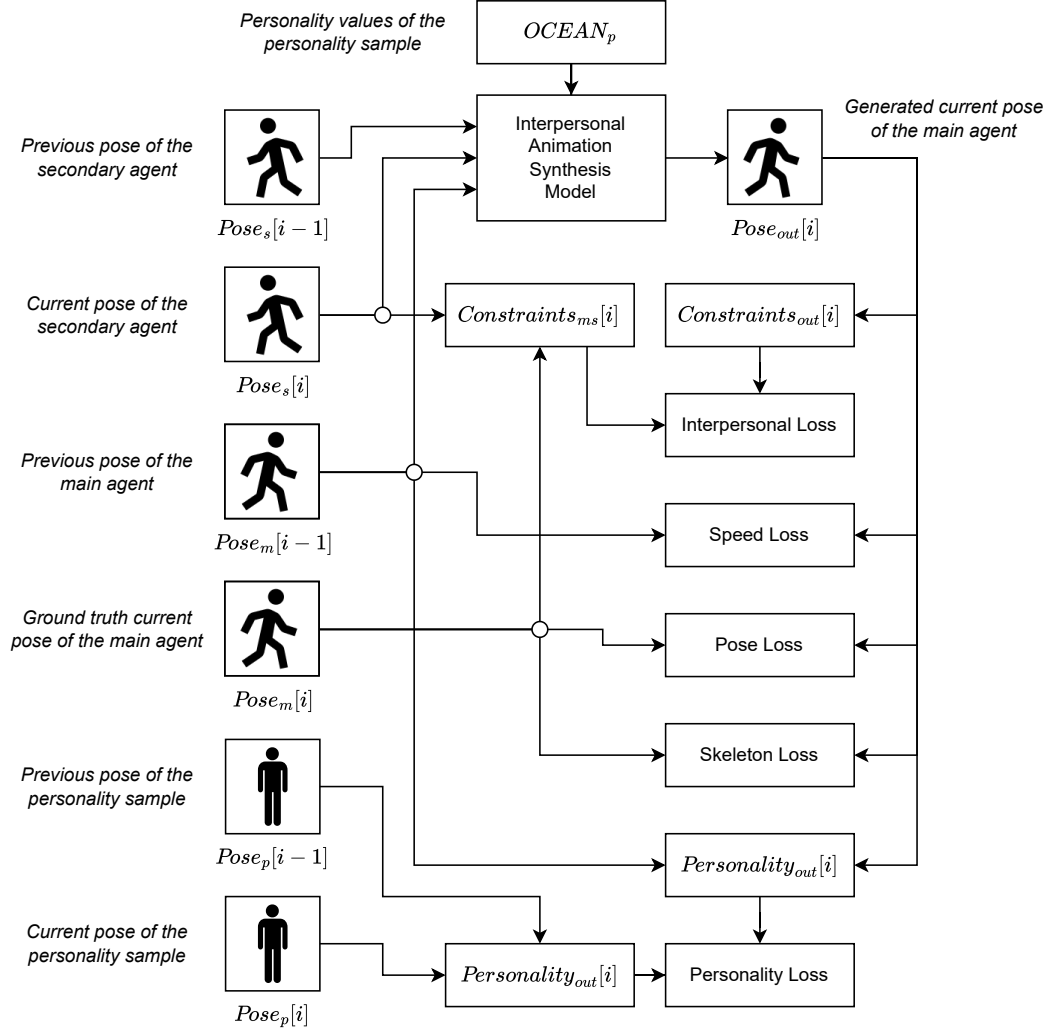


Figure 5.2: The data flow used in calculating the different losses of the model during training. We show the conjunction points with small circles.

but any attempt to include personality requires additional losses.

$$\mathcal{L}_{pose} = \frac{1}{N} \sum_{i=1}^N (Pose_{out}[i] - Pose_m[i])^2$$

- **Personality Loss ($\mathcal{L}_{personality}$):** This loss assures the generated pose for the main agent, $Pose_{out}[i]$, shows personality-related motion features $Personality_{out}[i]$, that is similar to $Personality_p[i]$. We calculate $Personality_{out}[i]$ over the generated pose $Pose_{out}[i]$ using the distances and angles that are shown to be influential on perceived personality. Similarly, $Personality_p[i]$ is calculated over $Pose_p[i]$. $Personality_x[i]$ includes parameters such as the Euclidean distance between the hands, the Euclidean distance between the right hand and the head, and the positive angle defined by the corner following the joints of the hips-spine-neck; the complete list of the personality motion features is included in the previous chapter. $\mathcal{L}_{pose}$ and $\mathcal{L}_{personality}$ can contradict; the optimization process finds a balance between the losses so that the generated output satisfies both conditions. However, this requires additional losses to ensure the human-likeness of the generated poses.

$$\mathcal{L}_{personality} = \frac{1}{N} \sum_{i=1}^N (Personality_p[i] - Personality_{out}[i])^2$$

- **Skeleton Loss ($\mathcal{L}_{skeleton}$):** We use global positions of the joints as $Pose_x$ features to simplify calculations during training. Most of our losses depend on the distances between joints, and conversion between different forms of joint representations (position and rotation) can introduce a bottleneck to the system’s performance. Furthermore, utilizing joint positions enables the network to consider different shapes of bodies automatically. For example, suppose the output pose should reach a certain target position. In that case, this is achieved directly using the distance-based losses, and the network outputs the location of the hand to satisfy all conditions. On the other hand, the generated positions for the main agent’s joints could result in a difference in limb lengths due to various losses, resulting in unnatural motion. $\mathcal{L}_{skeleton}$ assures that the generated pose $Pose_{out}[i]$ has limb lengths

that are the same as $Pose_m[i - 1]$. In other words, each limb’s length (the Euclidean distance between the linked joints) must be the same in the previous (ground truth) and current (generated) poses for the main agent. We calculate $Skeleton_x$, which is a vector of size 21 that contains the Euclidean distance between each linked joint based on the SMPL body definition, over $Pose_m[i - 1]$ and $Pose_{out}[i]$. $\mathcal{L}_{skeleton}$ is formed as the MSE between $Skeleton_m$ and $Skeleton_{out}$. Since the body configuration is time-invariant, we do not use time indices for the limb length vectors.

$$\mathcal{L}_{skeleton} = \frac{1}{N} \sum_{i=1}^N (Skeleton_m - Skeleton_{out})^2$$

- **Speed Loss ($\mathcal{L}_{speed}$):** The animation samples we utilize in training contain a pose at each frame, and since the motion capture devices use a constant frame rate, the resulting consecutive poses do not change dramatically. There is a maximum speed at which each joint can travel between the poses in two consecutive frames. $\mathcal{L}_{speed}$ assures the generated current pose of the main agent, $Pose_{out}[i]$, is not too different from the main agent’s previous pose, $Pose_m[i - 1]$. We use the MSE between $Pose_{out}[i]$ and $Pose_m[i - 1]$ to calculate the Speed Loss; this loss approaches the average joint speed of the samples during training.

$$\mathcal{L}_{speed} = \frac{1}{N} \sum_{i=1}^N (Pose_{out}[i] - Pose_m[i - 1])^2$$

- **Interpersonal Loss ($\mathcal{L}_{interpersonal}$):** Interpersonal dyadic actions follow specific rules that determine the motion of certain joints. Not all joints contribute to dyadic actions in the same way; the legs, for example, can be positioned freely during a handshake, but the hands have an essential role in such motions. Although $\mathcal{L}_{pose}$ directs the network to learn the output poses that complement the secondary agent’s motion, $\mathcal{L}_{interpersonal}$ adds additional importance to the joints forming interpersonal action. We calculate $Constraints_{ms}[i]$, the ground truth interpersonal constraints using the pair $\{Pose_m[i], Pose_s[i]\}$, and $Constraints_{out}[i]$, the interpersonal constraint features of the current output, using the pair $\{Pose_{out}[i], Pose_s[i]\}$. Both

calculations use the current pose of the secondary agent and utilize either the ground truth or the generated output for the main agent’s current pose. $Constraints_x[i]$ is a vector of features. The interpersonal constraint features include distances across the two agents’ joints, such as the distance between one agent’s left hand and the other agent’s left hand or one agent’s head and the other agent’s right-hand joints. These features are made from the possible pairs among the hand, elbow, foot, knee, head, and hip joints, such that each pair utilizes one joint of either the main or the secondary agent.

$$\mathcal{L}_{interpersonal} = \frac{1}{N} \sum_{i=1}^N (Constraints_{ms}[i] - Constraints_x[i])^2$$

The total loss $\mathcal{L}_{total}$ is calculated as the average of the sum of all loss elements:

$$\mathcal{L}_{total} = \frac{1}{5} (\mathcal{L}_{pose} + \mathcal{L}_{personality} + \mathcal{L}_{skeleton} + \mathcal{L}_{speed} + \mathcal{L}_{interpersonal}).$$

5.4 The Architecture

Our interpersonal animation synthesis network uses transformer encoders as building blocks. The network inputs the previous pose of the main agent $Pose_m[i-1]$, the previous and current poses of the secondary agent $Pose_s[i-1 : i]$, and the target personality traits $OCEAN_p$ to synthesize the main agent’s current pose expressing the desired traits $Pose_{out}[i]$. The input is fed to four sequential layers of transformer encoder units.

We utilize *TransformerEncoderLayer* of PyTorch, which is an implementation based on *Attention is All You Need* [148]. We used the default values for the dimension of the feedforward network model (2048), the dropout (0.1), and the ReLU (Rectified Linear Unit) activation function. A final fully connected layer decreased the output size to the number of output features (22 joints $\times$ three dimensions). We use the Rectified Adam [149] optimizer with its default values. We track the sum of all losses and save the network with the lowest loss. We train on an RTX 4060 GPU for 500 epochs with a batch size of 32.

5.5 Evaluation

The evaluation of the generative system consists of generating the stimuli, performing the online study, and reporting the results. We discuss the outcomes regarding the system’s performance for each personality factor.

5.5.1 The Stimuli

We evaluate our system based on how well it can distinguish opposing factors. To this end, we chose two handshake and two high-five animations from the Inter-X dataset, with the main agent’s ground truth personality labels being close to neutral. In other words, the sample with the smallest average distance to a hypothetical average personality where all factors are 0, assuming the personality factor values are normalized to the range $[-1, 1]$. Since the dataset contains multiple takes for the same subjects interacting, we choose the take with fewer distracting elements, such as additional movements not part of the main action. We use our generative model to synthesize altered versions of each motion using different target personalities, low and high. An example output where we express high extraversion is shown in Figure 5.3, where the body’s reach is extended.

For each original sample, HS_1 and HS_2 for handshaking and HF_1 and HF_2 for high-five animations, we generate an altered version such as $X_{HS_1}^y$, where X refers to the selected factor to alter as one of $\{O, C, E, A, N\}$, and y refers to the direction of change as one of $\{+, -\}$. For example, $E_{HS_1}^+$ refers to the generated animation using the secondary agent motion from the HS_1 sample, focusing on expressing high extraversion. Consequently, we expect this sample’s perceived extraversion to be high. In contrast, the perceived extraversion of the generated $E_{HS_1}^-$ is expected to be low. Our user study compares the low and high versions of the same animation, focusing on individual traits one at a time.

We compare each task’s opposing polarity versions of the same animation based on a single factor. For example, the success in distinguishing extraversion

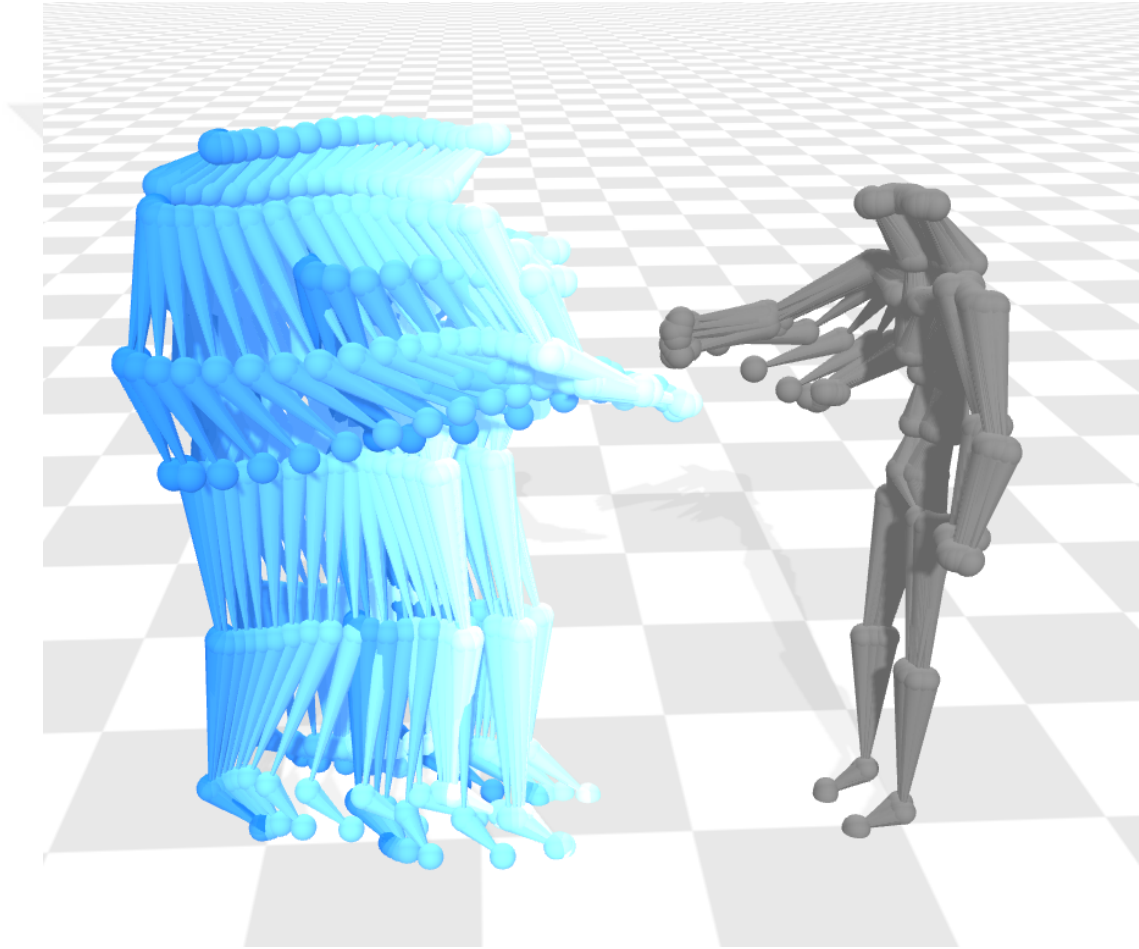


Figure 5.3: A sample generated animation rendered using stick figures. The gray figures depict the secondary agent’s movements, and the blue figures show the generated main agent’s motion. We overlay the agent poses from multiple frames to depict the movement; the color of the main agent shifts to white in later frames.

for the sample HS_1 is based on the users' preference of $E_{HS_1}^+$ over $E_{HS_1}^-$ for the high trait perception. Similarly, we expect the user to choose $E_{HS_1}^-$ over $E_{HS_1}^+$ for the question that measures the low trait perception.

5.5.2 The User Study

Our survey questions utilize the Ten Item Personality Inventory (TIPI) [127], where each question measures a specific factor of FFM either with direct or inverse relation. The users click on the buttons to give their answer as either *Left*, *Equal*, or *Right* to the questions shown in Table 5.1. We randomize the left-right arrangement of the samples in each task to avoid any bias due to placement.

No	Factor	Polarity	Question
1	E	+	Which is more extraverted, enthusiastic?
2	A	−	Which is more critical, quarrelsome?
3	C	+	Which is more dependable, self-disciplined?
4	N	+	Which is more anxious, easily upset?
5	O	+	Which is more open to new experiences, complex?
6	E	−	Which is more reserved, quiet?
7	A	+	Which is more sympathetic, warm?
8	C	−	Which is more disorganized, careless?
9	N	−	Which is calmer, emotionally stable?
10	O	−	Which is more conventional, uncreative?

Table 5.1: The questions asked in the user study to evaluate each generated pair. The participants answer each question as one of *Left*, *Equal*, or *Right*.

The participants view the two animations side by side. The animations loop unless the user clicks on the pause button. Users can play the generated samples one by one or in a synchronized manner. The animations are rendered using stick figures to avoid bias due to the body shape. The participants view the agents from a fixed angle. We did not give any camera controls to the user so that each participant could view the scene from the same angle. Otherwise, a difference in users' viewing angle of choice could lead to different perceptions [150]. The participants were required to answer each question to submit the task. They were free to take a break and resume the experiment. The participants were allowed

to change their answers for the current task, which is fixed after submitting the task, with no return allowed to a previously submitted task. Figure 5.4 shows an example view of the participants.

We perform a user study ¹ to evaluate the results. We included 20 tasks (four ground truth samples transformed with five target factors, high and low) for each participant. We acquired participants using the online crowd-sourcing platform Prolific. Upon completing all the tasks, the participants were paid based on their average time spent on the study. Thirty individuals (16 male, 11 female, three not specified) participated in our evaluation study with a mean age of 34.16 ± 12.73 . Each participant spent an average of 18.13 ± 4.26 minutes on the 20 tasks completed using our online survey page. We randomized the order in which the sample pairs appear to prevent bias. The participants were from different countries: 12 from South Africa, nine from Europe, five from the United States, and the rest from other continents.

5.5.3 Results

For each generated sample pair, we collect data on the perceived personality of all factors. We expect the highest difference regarding the target factor where the network alters; however, it is common that personality factors are perceived in relation to each other [15]. This behavior is especially true when there are limited cues on certain factors. The TIPI questionnaire includes two questions for each personality factor, one reverse-scored. To analyze the results, we replace user choices of left and right with high and low, based on the placement of the samples. For example, the expected answer for a question measuring positive extraversion is the high sample (which can be on the right or the left based on the random placement). Similarly, we expect the low version to be selected for a reverse-scored question. We map user choices to 1 if they select the expected answer, 0 if they choose equal, and -1 if they choose the opposite sample. Then, we sum up the questions that focus on the same factor.

¹Bilkent University Ethical Committee for Human Research approved the user study with the decision number IAEK.2024.07.16.02.

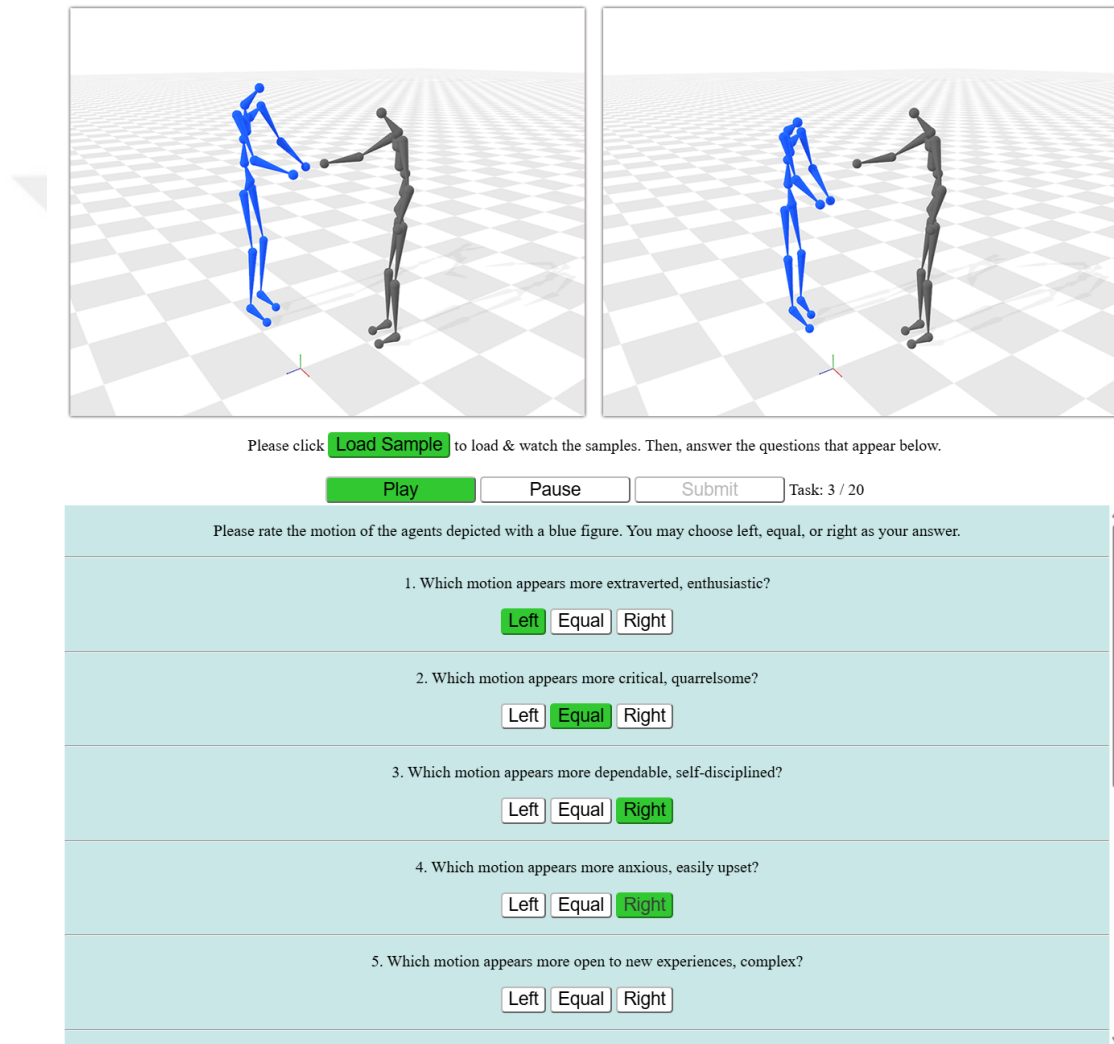


Figure 5.4: An example screenshot from the user study for the evaluation. The samples are shown side-by-side in a loop. The participant can pause and resume the generated animations using the buttons. Each question should be answered before submitting the task; the submit button is enabled once the user chooses an answer for each question.

For instance, if the participant chose a “high” sample for the question asking “Which is more extraverted, enthusiastic?”, we get a score of 1 for extraversion. If for the reverse-scored question for extraversion, asking “Which is more reserved, quiet?”, they choose the “low” sample, then we get another score of 1 for extraversion, summing up to 2. Consequently, for each personality factor, we obtain a value in the range $[-2, 2]$ that shows the confidence of the user’s answer in the system’s performance. The score of 2 corresponds to the system successfully distinguishing the target factor in the correct direction. The score of -2 suggests that the system creates a difference, but in the opposite direction, which can still be useful if we reverse the signs. On the other hand, a score near 0 means that the system cannot create an observable difference. For convenience, we normalize these measurements to the $[-1, 1]$ range.

To measure the significance of the difference between the paired samples, we compare the participants’ answers to the questions in contrast to uniform randomness. For example, the task where we compare $E_{HS_1}^+$ and $E_{HS_1}^-$ is named as E_{HS_1} ; this is the first handshaking animation altered for high and low extraversion expression. Based on 30 participants, the expected uniform random answers for each personality question in this task would be 10 *Left*, 10 *Equal*, and 10 *Right*. We use the T-test to compare the actual distribution to the uniform randomness, and the value we obtain shows the significance of the difference. If the obtained p-value is close to zero, the mean of the distribution is significantly different than the random distribution.

We report the T-test results for each task and each question measuring a personality factor’s polarity in Table 5.2. In this table, we also report the combined overall results for the tasks that focus on altering the same personality factor. For example, E_{Comb} gathers all four tasks that focus on altering extraversion (E_{HS_1} , E_{HS_2} , E_{HF_1} , and E_{HF_2}) together. In this case, the distribution is compared to the uniform randomness with 120 participants; since the number of responses is increased, having a more significant result is more likely. Consequently, we ignore the significant combined results if none of the individual tasks are significant. For example, the question focusing on *O+* has a significant result for E_{Comb} , but the individual results have no significant effect; therefore, we conclude that the

differences due to extraversion have no significant effect on openness perception.

We observe a strong effect on personality perception when the network focuses on altering extraversion and agreeableness. The differences due to openness have no significant effect, and the differences due to conscientiousness weakly influence extraversion perception. The extraversion differences of the output have a major effect on extraversion perception and a slight effect on neuroticism perception. The changes due to agreeableness mainly influence agreeableness perception but also slightly affect extraversion perception. Neuroticism’s effect is very limited in perception; the samples that differ in neuroticism are perceived differently regarding the extraversion factor.

Table 5.3 shows the results when we combine the questions that measure a particular factor using the standard TIPI evaluation. The reverse-scored factor answers are taken with the opposite sign in this case. For example, if the user chose the “high” sample for $E+$, represented with $+1$, and the “low” sample for $E-$, represented with -1 , then we calculate the overall E score as $+1 - (-1) = 2$. We normalize the sum back to the $[-1, 1]$ range to report the values with the T-test results, in contrast to uniform randomness and the mean of the distribution. We expect the mean score of the personality measurements to be close to 1 for the factor that the network aims to alter in a successful system. For example, if the changes done by the system regarding extraversion create an observable difference, the extraversion mean for this group of samples should be close to 1. We also observe that the changes made by the network regarding a particular factor also affect the perception of other factors. In particular cases, this effect could be in the reverse direction. For example, we observe that the extraversion differences of the network create an inverse impact on agreeableness perception. The samples with high extraversion are perceived as low in agreeableness.

Combining the user answers to the questions measuring the same factor gives a clearer summary of the network’s effect. Differences in conscientiousness made by the network slightly affect extraversion perception. The network’s high conscientiousness is perceived as high extraversion. The network’s effect on extraversion is translated into a significant difference for extraversion perception, with a lower

Set	O+	O−	C+	C−	E+	E−	A+	A−	N+	N−
O_{HS_1}	.878	.499	.744	.865	.512	.628	.321	.744	.390	.869
O_{HS_2}	.371	.744	.872	.720	.252	.194	.523	.001	.628	.872
O_{HF_1}	.252	.307	.348	.277	.081	.307	.265	.335	.158	.206
O_{HF_2}	.869	.628	.277	.872	.544	.646	.646	.875	.348	.646
$O_{Comb.}$	.937	.450	.422	.355	.036	.101	.106	.685	.060	.468
C_{HS_1}	.001	.628	.289	.762	.041	.158	.429	.224	.597	.729
C_{HS_2}	.628	.875	.762	.277	.110	.637	.744	.628	.110	.534
C_{HF_1}	.750	.875	.452	.206	.005	.065	.110	.100	.534	.637
C_{HF_2}	.646	.534	.429	.441	.348	.646	.073	.881	.441	.238
$C_{Comb.}$	.747	.689	.141	.159	.001	.111	.023	.571	.194	.322
E_{HS_1}	.206	.441	.300	.088	.001	.002	.653	.218	.096	.544
E_{HS_2}	.136	.757	.553	.884	.001	.001	.452	.452	.046	.081
E_{HF_1}	.088	.462	.646	.881	.001	.001	.277	.768	.206	.348
E_{HF_2}	.878	.744	.147	.125	.051	.019	.472	.096	.033	.125
$E_{Comb.}$	.032	.434	.075	.120	.001	.001	.129	.047	.001	.016
A_{HS_1}	.071	.441	.300	.088	.179	.065	.005	.129	.307	.544
A_{HS_2}	.194	.224	.584	.597	.300	.250	.002	.004	.065	.277
A_{HF_1}	.597	.241	.512	.646	.762	.872	.194	.762	.637	.348
A_{HF_2}	.429	.750	.100	.523	.024	.206	.002	.032	.158	.194
$A_{Comb.}$	.075	.497	.138	.088	.014	.026	.001	.001	.055	.307
N_{HS_1}	.417	.869	.265	.750	.081	.252	.869	.523	.158	.348
N_{HS_2}	.860	.720	.390	.404	.021	.534	.737	.860	.512	.417
N_{HF_1}	.001	.512	.744	.045	.073	.136	.872	.417	.091	.757
N_{HF_2}	.486	.628	.206	.878	.158	.534	.206	.182	.147	.348
$N_{Comb.}$	.001	.396	.138	.104	.001	.052	.322	.403	.008	.129

Table 5.2: The pairwise comparisons focus on altering each personality factor. The columns indicate the questions measuring each polarity of personality factors. Rows indicate either a specific altered motion pair or the combined version of all such pairs, focusing on altering the same factor. We report the p-values of the two-tailed T-test results in contrast to uniform randomness. The statistically significant results where $p > 0.05$ are highlighted.

Set	ρ_O	μ_O	ρ_C	μ_C	ρ_E	μ_E	ρ_A	μ_A	ρ_N	μ_N
O_{HS_1}	.635	.083	.927	-.017	.509	.117	.720	.067	.706	.067
O_{HS_2}	.723	-.067	.922	.017	.168	.250	.716	.067	.854	.033
O_{HF_1}	.218	.217	.255	.217	.112	.283	.247	.217	.138	.283
O_{HF_2}	.716	-.067	.468	.133	.556	.117	.729	.067	.407	.150
$O_{Comb.}$	.646	.042	.332	.087	.034	.192	.247	.104	.123	.133
C_{HS_1}	.782	-.050	.653	.083	.045	.367	.850	-.033	.622	-.083
C_{HS_2}	.711	.067	.431	.150	.227	.217	.653	.083	.197	.233
C_{HF_1}	.931	-.017	.262	.217	.009	.450	.064	.333	.531	.117
C_{HF_2}	.928	-.017	.390	.167	.780	.050	.241	.200	.259	.200
$C_{Comb.}$	.963	-.004	.099	.154	.003	.271	.097	.146	.190	.117
E_{HS_1}	.258	.217	.108	.300	.001	.700	.337	-.183	.179	.250
E_{HS_2}	.313	.183	.800	.050	.001	.783	.386	-.167	.042	.400
E_{HF_1}	.145	.267	.731	.067	.001	.750	.445	-.150	.201	.233
E_{HF_2}	.784	-.050	.090	.300	.013	.450	.169	-.267	.044	.350
$E_{Comb.}$	.103	.154	.054	.179	.001	.671	.039	-.192	.001	.308
A_{HS_1}	.141	.283	.087	.300	.069	-.333	.020	.450	.378	-.167
A_{HS_2}	.168	.250	.001	.000	.207	-.250	.001	.583	.100	-.317
A_{HF_1}	.376	-.150	.544	.117	.786	-.050	.354	.167	.390	.150
A_{HF_2}	.513	.117	.205	.233	.044	-.367	.004	.550	.119	-.283
$A_{Comb.}$	.150	.125	.064	.163	.007	-.250	.001	.438	.092	-.154
N_{HS_1}	.708	-.067	.423	-.150	.105	.300	.784	.050	.190	.250
N_{HS_2}	.774	.050	.350	-.167	.090	.300	.774	-.050	.411	.150
N_{HF_1}	.716	.067	.345	-.167	.067	.333	.597	-.100	.276	.200
N_{HF_2}	.528	.117	.419	-.150	.247	.217	.120	-.267	.175	.250
$N_{Comb.}$	.632	.042	.075	-.158	.002	.287	.296	-.092	.018	.212

Table 5.3: The results when we combine the measurements that focus on the same personality factor. The user’s answer is summed with the reversed inverse polarity question for each question that focuses on a specific factor. We report the measurements’ significance (ρ) and mean (μ). A mean of 1 indicates that the majority chooses the sample that expresses the high polarity to represent the high perception, which is what we expect from a successful system. Columns depict personality factors, and the rows show the tasks. We report the p-values of the two-tailed T-test results in contrast to uniform randomness. We highlight statistically significant results with $p > 0.05$.

impact on neuroticism and a slightly reversed effect on agreeableness. The high extraversion output of the model is perceived as high in extraversion, moderately high in neuroticism, and slightly low in agreeableness. The network’s effect on agreeableness moderately impacts the agreeableness perception, and the extraversion perception is slightly reversed. The high agreeableness output of the network is perceived as moderately high in agreeableness and slightly low in extraversion. The network’s effect on neuroticism is limited; the high neuroticism output is perceived as slightly high in extraversion and neuroticism.

Figure 5.5 shows the number of participants who chose the *Low*, *Equal*, and *High* answers for each original motion and personality factor combination. We report the numbers for each TIPI question, focusing on a factor’s high or low polarity. We observe that, in general, the network’s effect on the difference between original motions is similar. We expect the factor that the network focuses on altering to have more *High* answers for the direct scored questions (+), and more *Low* answers for the reverse scored questions (−).

Figure 5.6 depicts the results when the combined tasks utilize different motions. To this end, disregard the different original motions and group the samples based on the factor the network focuses on to alter. For example, all four tasks that focus on altering extraversion (E_{HS_1} , E_{HS_2} , E_{HF_1} , and E_{HF_2}) are combined as $E_{Comb.}$. We observe that the highest perceived difference is due to the extraversion changes done by the network, followed by agreeableness.

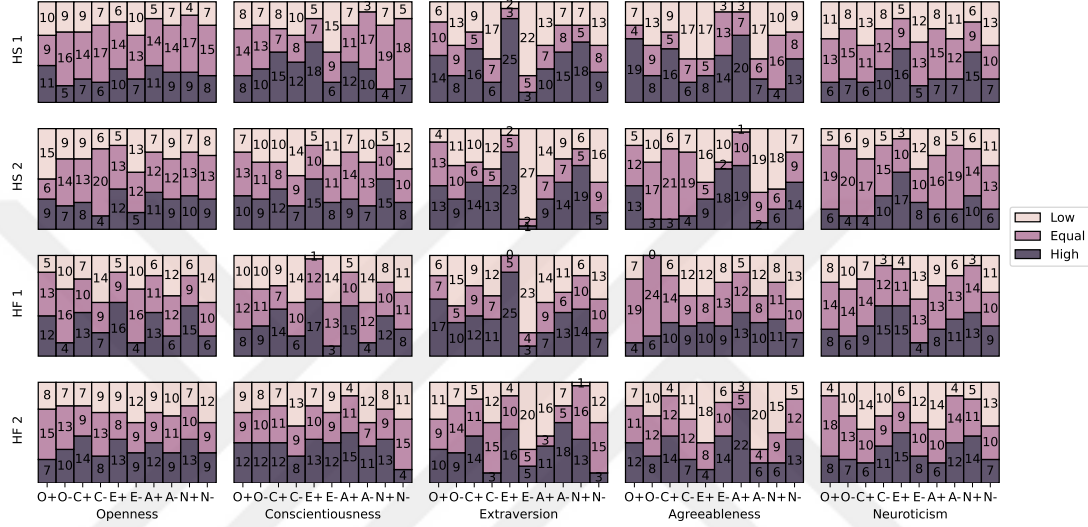


Figure 5.5: Bar plots of the users' answers to the questions for each original motion (rows) and each personality factor focused by the network (columns). The values depict the number of participants who chose a specific version.

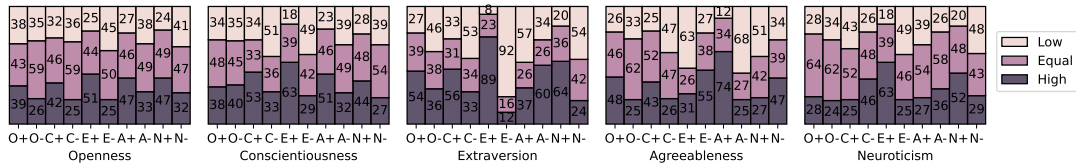


Figure 5.6: Bar plots of the users' answers to the questions for each personality factor focused by the network, combining the different original motions. The values depict the number of participants who chose a specific version.

Chapter 6

Conclusions and Future Research Directions

This thesis work focuses on personality expression in target-sensitive interpersonal animations. We first focus on identifying the effective motion parameters in personality estimation and expression for a generalizable solution. Motion parameters enable replicating various personality styles without changing the semantic action categories. As motion datasets with personality labels are scarce, we utilize crowdsourcing to introduce personality annotations to a subset of single-agent animations from well-known datasets. We select a subset of motion features that influence apparent personality traits using feature importance analysis on the personality annotations.

Our interpersonal animation generation system uses two types of motion for training. The dyadic motion samples are used to train the semantic context of animation. Our model utilizes the global joint positions for target-sensitive motion synthesis; as a result, interpersonal animations such as handshaking can be generated without losing the context of interaction. However, the use of joint positions may negatively impact body integrity. We utilize skeletal loss to enforce the initial measurements of the limbs. We use dyadic animations and single-agent motion samples to introduce personality to our model. Since the existing dyadic

action datasets lack style variance, using single-agent samples with high style variance helps improve the expressiveness of the results. Our transformer-based model inputs the current and previous poses of the secondary agent, the previous pose of the main agent, and the desired personality traits to generate the current pose of the main agent.

We perform a user study to evaluate various dyadic animations expressing different personality traits. The results show that the system can significantly differ in perceived extraversion and agreeableness; however, the system’s effect on openness, conscientiousness, and neuroticism is limited. Openness is generally considered a difficult factor to express in short sequences. The lack of speed variance in the system’s output could be the reason for the reduced effect on conscientiousness and neuroticism, which are closely related to speed and acceleration features. We synthesize the main agent’s current motion, relying on the secondary agent’s ground truth previous and current poses, which limits the generated poses. The personality-based variance is learned through the motion parameters, which cause it to be less exaggerated. Consequently, the lack of variance in certain types of motion features, like animation speed, in the dyadic samples translates into a similar lack of variance in the generated samples. Future work can utilize personality-based augmentations to improve the variance in speed and acceleration-related parameters.

We focus on single-trait differences, leaving the evaluation of the effect of utilizing a combination of factors. For example, expressing the same polarity for multiple personality factors is expected to have an additive effect; however, factors can have an interesting interaction when some express high polarity while some reflect low polarity. Although generating such combinations is supported by our current model, we leave the evaluation of such output to future studies, as the variations result in too many combinations. A follow-up study that focuses on this subject should consider the effect of all 3^5 combinations; that is, low, neutral, or high traits for each of the five personality factors. Additionally, the magnitude of the target personality factors can be used on a larger scale. Currently, we limit the personality input of the network to the extremes and the neutral to observe the highest effect possible; when combining multiple factor expressions,

the magnitude of the difference in each factor may need further adjustments.

We focused on simpler dyadic actions such as high-fiving and handshaking. Future work can include more complex categories, such as hugging. One major issue for such animations is that the degree of freedom in such actions is hard to estimate without a rigid body representation from skeletal data. For example, hugging mostly involves contact between the hands and the body, but the joint representations do not support this information. More complex data can be used in training, which also considers the body’s surface area, such that the network can learn the contact points in finer detail. Our system only operates on joint positions; however, a volumetric representation can enable higher-quality results. Rendering the animations using stick figures is more forgiving, as any small intersection between the joints is not very pronounced; in contrast, using realistic rendering would negatively impact the plausibility. For example, currently, we do not focus on finger animations; therefore, the finger joints are omitted. Any realistic rendering applied to our results should also consider the finger positions.

Our interpersonal animation generation system focused on body motion; consequently, we did not include facial expressions or dialogue in our evaluated outputs. Future studies can examine the interaction between expressive interpersonal animations and different facial expressions and dialogue choices. For example, while smiling positively affects personality perception, its influence can change based on the context of the dyadic action. Smiling while handshaking can have a different effect than smiling while pushing someone. Testing such facial expressions also requires realistic rendering, which could affect the perceived quality of motion. For example, when rendered with simpler shapes, certain movement artifacts do not drastically impact the perceived realism; however, due to an uncanny effect, the small artifacts on the generated motion could negatively impact the realism when rendered on a realistic figure.

Rather than providing a comprehensive solution to apply personality to general motion, we examined the animation elements that communicate different personality traits in the context of interpersonal actions. The neural models trained for this thesis work are fine-tuned for specific datasets and, therefore, cannot

be used directly with different representations. Since animation standards can vary dramatically, using any motion recorded using a different setup or format requires conversions. Future work can also pursue a unified framework for adjusting generic human animations, such that regardless of the format and size differences, the high-level parameters can be controlled seamlessly. For instance, such a system could create an expanding posture for the bodies of different skeletal configurations and sizes, training on semantic categories.

We hope this thesis work and results inspire future work that aims for more enriched animations. Personality offers one way to control the expressive nature of human motion; however, other psychological concepts like emotions and mood also offer similar tools, and a final future research direction is to analyze the interaction between personality, emotion, and mood in the context of target-sensitive interpersonal animations.

Bibliography

- [1] I. Zwir, J. Arnedo, C. Del-Val, L. Pulkki-Råback, B. Konte, S. S. Yang, R. Romero-Zaliz, M. Hintsanen, K. M. Cloninger, D. Garcia, *et al.*, “Uncovering the complex genetics of human character,” *Molecular Psychiatry*, vol. 25, no. 10, pp. 2295–2312, 2020.
- [2] M. Briffa and A. Weiss, “Animal personality,” *Current Biology*, vol. 20, no. 21, pp. R912–R914, 2010.
- [3] S. D. Gosling and O. P. John, “Personality dimensions in nonhuman animals: A cross-species review,” *Current Directions in Psychological Science*, vol. 8, no. 3, pp. 69–75, 1999.
- [4] F. B. De Waal, “What is an animal emotion?,” *Annals of the New York Academy of Sciences*, vol. 1224, no. 1, pp. 191–206, 2011.
- [5] G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. R. D’Orazio, “Human gait analysis in neurodegenerative diseases: A review,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 229–242, 2021.
- [6] J. Sun, P. Wu, Y. Shen, Z. Yang, H. Li, Y. Liu, T. Zhu, L. Li, K. Zhang, and M. Chen, “Relationship between personality and gait: Predicting personality with gait features,” in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine*, BIBM ’18, (Piscataway, NJ, USA), pp. 1227–1231, IEEE, 2018.

- [7] Y. Wen, B. Li, D. Chen, and T. Zhu, “Reliability and validity analysis of personality assessment model based on gait video,” *Frontiers in Behavioral Neuroscience*, vol. 16, 2022. Article no. 901568, 13 pages.
- [8] Y. Wang, J. E. F. Tree, M. Walker, and M. Neff, “Assessing the impact of hand motion on virtual character personality,” *ACM Transactions on Applied Perception*, vol. 13, no. 2, 2016. Article no. 9, 23 pages.
- [9] R. Subramanian, Y. Yan, J. Staiano, O. Lanz, and N. Sebe, “On the relationship between head pose, social attention and personality prediction for unstructured and dynamic group interactions,” in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI ’13, (New York, NY, USA), pp. 3–10, ACM, 2013.
- [10] S. J. Guy, S. Kim, M. C. Lin, and D. Manocha, “Simulating heterogeneous crowd behaviors using personality trait theory,” in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA ’11, (New York, NY, USA), pp. 43–52, ACM, 2011.
- [11] P. Wagner, Z. Malisz, and S. Kopp, “Gesture and speech in interaction: An overview,” *Speech Communication*, vol. 57, pp. 209–232, 2014.
- [12] K. Crockett, J. O’Shea, and W. Khan, “Automated deception detection of males and females from non-verbal facial micro-gestures,” in *Proceedings of the International Joint Conference on Neural Networks*, IJCNN ’20, (Piscataway, NJ, USA), pp. 1–7, IEEE, 2020.
- [13] H. Chen, H. Shi, X. Liu, X. Li, and G. Zhao, “SMG: A micro-gesture dataset towards spontaneous body gestures for emotional stress state analysis,” *International Journal of Computer Vision*, vol. 131, no. 6, pp. 1346–1366, 2023.
- [14] N. Dael, M. Mortillaro, and K. R. Scherer, “Emotion expression in body action and posture,” *Emotion*, vol. 12, no. 5, pp. 1085–1101, 2012.
- [15] H. J. Smith and M. Neff, “Understanding the impact of animated gesture performance on personality perceptions,” *ACM Transactions on Graphics*, vol. 36, no. 4, 2017. Article no. 49, 12 pages.

- [16] M. A. Akber, T. Ferdousi, R. Ahmed, R. Asfara, R. Rab, and U. Zakia, “Personality and emotion—a comprehensive analysis using contextual text embeddings,” *Natural Language Processing Journal*, vol. 9, 2024. Article no. 100105, 13 pages.
- [17] X. Zhao and N. I. Badler, “Interactive body awareness,” *Computer-Aided Design*, vol. 26, no. 12, pp. 861–867, 1994.
- [18] L. S. Scimmi, M. Melchiorre, S. Mauro, S. Pastorelli, *et al.*, “Multiple collision avoidance between human limbs and robot links algorithm in collaborative task,” in *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics*, ICINCO ’18, pp. 301–308, 2018.
- [19] M.-C. Lee, M. Trinh, and Z. Deng, “Multimodal turn analysis and prediction for multi-party conversations,” in *Proceedings of the 25th International Conference on Multimodal Interaction*, ICMI ’23, pp. 436–444, 2023.
- [20] H. Liang, W. Zhang, W. Li, J. Yu, and L. Xu, “InterGen: Diffusion-based multi-human motion generation under complex interactions,” *International Journal of Computer Vision*, vol. 132, pp. 3463–3483, 2024.
- [21] L. Xu, X. Lv, Y. Yan, X. Jin, S. Wu, C. Xu, Y. Liu, Y. Zhou, F. Rao, X. Sheng, *et al.*, “Inter-X: Towards versatile human-human interaction analysis,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR ’24, (Piscataway, NJ, USA), pp. 22260–22271, IEEE, 2024.
- [22] M. Kobayashi, C. Liao, K. Inoue, S. Yojima, and M. Takahashi, “Motion capture dataset for practical use of AI-based motion editing and stylization,” *CoRR*, vol. abs/2306.08861, 2023. Available at <https://doi.org/10.48550/arXiv.2306.08861>, Accessed: 2025/08/11.
- [23] S. Ghorbani, Y. Ferstl, D. Holden, N. F. Troje, and M.-A. Carbonneau, “ZeroEGGS: Zero-shot Example-based Gesture Generation from Speech,” *Computer Graphics Forum*, vol. 42, no. 1, pp. 206–216, 2023.

- [24] J. M. Digman, “Personality structure: Emergence of the five-factor model,” *Annual Review of Psychology*, vol. 41, no. 1, pp. 417–440, 1990.
- [25] Z. Erkoç, S. Demirci, S. Sonlu, and U. Gdkbay, “Skeleton-based personality recognition using laban movement analysis,” in *Understanding Social Behavior in Dyadic and Small Group Interactions*, pp. 74–87, PMLR, 2022.
- [26] H. Mailee, S. Sonlu, and U. Gdkbay, “Personality expression in cartoon animal characters using sasang typology,” *Computer Animation and Virtual Worlds*, vol. 34, no. 3-4, p. e2164, 2023.
- [27] S. Sonlu, K. Shabanovi, U. Gdkbay, and F. Durupinar, “Emotional expression as a means of communicating virtual human personalities,” in *Modeling Visual Aesthetics, Emotion, and Artistic Style*, pp. 293–312, Springer, 2023.
- [28] S. Sonlu, Y. Doėan, A. . Ergzen, M. E. nalan, S. Demirci, F. Durupinar, and U. Gdkbay, “Towards understanding personality expression via body motion,” in *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 628–631, IEEE, 2024.
- [29] A. Yurtoėlu, S. Sonlu, Y. Doėan, and U. Gdkbay, “Personality perception in human videos altered by motion transfer networks,” *Computers & Graphics*, vol. 119, 2024. Article no. 103886, 11 pages.
- [30] S. Sonlu, Y. Doėan, A. . Ergzen, M. E. nalan, S. Demirci, F. Durupinar, and U. Gdkbay, “Human movement personality detection parameters,” in *Proceedings of the 32nd Signal Processing and Communications Applications Conference, SIU '24*, pp. 1–4, IEEE, 2024.
- [31] S. Sonlu, H. . Demir, and U. Gdkbay, “Personality expression using co-speech gesture,” *ACM Transactions on Applied Perception*, vol. 22, no. 2, 2025. Article no. 7, 20 pages.
- [32] S. Sonlu, B. Bendiksen, F. Durupinar, and U. Gdkbay, “Effects of embodiment and personality in LLM-based conversational agents,” in *Proceedings of the IEEE Conference Virtual Reality and 3D User Interfaces, VR '25*, (Piscataway, NJ, USA), pp. 718–728, IEEE, 2025.

- [33] M. Efe Sak, S. Sonlu, and U. Gdkbay, “Talk with Socrates: Relation between perceived agent personality and user personality in LLM-based natural language dialogue using virtual reality,” *Computer Animation and Virtual Worlds*, vol. 36, no. 3, 2025. Article no. e70033, 11 pages.
- [34] F. Multon, L. France, M.-P. Cani-Gascuel, and G. Debunne, “Computer animation of human walking: a survey,” *The Journal of Visualization and Computer Animation*, vol. 10, no. 1, pp. 39–54, 1999.
- [35] L. Herda, P. Fua, R. Plankers, R. Boulic, and D. Thalmann, “Skeleton-based motion capture for robust reconstruction of human motion,” in *Proceedings of Computer Animation*, CA ’00, (Piscataway, NJ, USA), pp. 77–83, IEEE, 2000.
- [36] M. Yahya, J. A. Shah, K. A. Kadir, Z. M. Yusof, S. Khan, and A. Warsi, “Motion capture sensing techniques used in human upper limb motion: A review,” *Sensor Review*, vol. 39, no. 4, pp. 504–511, 2019.
- [37] S. Vafadar, W. Skalli, A. Bonnet-Lebrun, A. Assi, and L. Gajny, “Assessment of a novel deep learning-based marker-less motion capture system for gait study,” *Gait & Posture*, vol. 94, pp. 138–143, 2022.
- [38] A. Shafaei and J. J. Little, “Real-time human motion capture with multiple depth cameras,” in *Proceedings of the 13th Conference on Computer and Robot Vision*, CRV ’16, (Piscataway, NJ, USA), pp. 24–31, IEEE, 2016.
- [39] Y. Zhang, Z. Li, L. An, M. Li, T. Yu, and Y. Liu, “Lightweight multi-person total motion capture using sparse multi-view cameras,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, CVPR ’21, (Piscataway, NJ, USA), pp. 5560–5569, IEEE, 2021.
- [40] J. Li, X. Liu, Z. Wang, H. Zhao, T. Zhang, S. Qiu, X. Zhou, H. Cai, R. Ni, and A. Cangelosi, “Real-time human motion capture based on wearable inertial sensor networks,” *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8953–8966, 2021.

- [41] A. Aristidou, D. Cohen-Or, J. K. Hodgins, and A. Shamir, “Self-similarity analysis for motion capture cleaning,” *Computer Graphics Forum*, vol. 37, no. 2, pp. 297–309, 2018.
- [42] S. A. Skogstad, “Filtering motion capture data for real-time applications,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME ’13, pp. 142–147, 2013.
- [43] D. Holden, “Robust solving of optical motion capture data by denoising,” *ACM Transactions on Graphics*, vol. 37, no. 4, 2018. Article no. 165, 12 pages.
- [44] A. Bruderlin, C. G. Teo, and T. Calvert, “Procedural movement for articulated figure animation,” *Computers & Graphics*, vol. 18, no. 4, pp. 453–461, 1994.
- [45] Y.-H. Lin, C.-Y. Liu, H.-W. Lee, S.-L. Huang, and T.-Y. Li, “Evaluating emotive character animations created with procedural animation,” in *Intelligent Virtual Agents: 9th International Conference, IVA 2009 Amsterdam, The Netherlands, September 14-16, 2009 Proceedings 9*, (Berlin, Heidelberg, Germany), pp. 308–315, Springer, 2009.
- [46] M. Okaya and T. Takahashi, “Human relationship modeling in agent-based crowd evacuation simulation,” in *International Conference on Principles and Practice of Multi-Agent Systems*, (Berlin, Heidelberg, Germany), pp. 496–507, Springer, 2011.
- [47] M. Xu, C. Li, P. Lv, W. Chen, Z. Deng, B. Zhou, and D. Manocha, “Emotion-based crowd simulation model based on physical strength consumption for emergency scenarios,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 6977–6991, 2020.
- [48] G. Rockenbach, C. Boeira, D. Schaffer, A. Antonitsch, and S. R. Musse, “Simulating crowd evacuation: From comfort to panic situations,” in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, IVA ’18, (New York, NY, USA), pp. 295–300, Association for Computing Machinery, 2018.

- [49] T. Geller, “Overcoming the uncanny valley,” *IEEE Computer Graphics and Applications*, vol. 28, no. 4, pp. 11–17, 2008.
- [50] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. Te Boekhorst, and K. L. Koay, “Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion,” *Autonomous Robots*, vol. 24, pp. 159–178, 2008.
- [51] S. Thomas, Y. Ferstl, R. McDonnell, and C. Ennis, “Investigating how speech and animation realism influence the perceived personality of virtual characters and agents,” in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*, VR ’22, (Piscataway, NJ, USA), pp. 11–20, IEEE, 2022.
- [52] T.-Y. Lee, C.-H. Lin, Y.-S. Wang, and T.-G. Chen, “Animation key-frame extraction and simplification using deformation analysis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 478–486, 2008.
- [53] J. Ojeda, J. Martínez-Reina, and J. Mayo, “The effect of kinematic constraints in the inverse dynamics problem in biomechanics,” *Multibody System Dynamics*, vol. 37, pp. 291–309, 2016.
- [54] W. Yoshizaki, Y. Sugiura, A. C. Chiou, S. Hashimoto, M. Inami, T. Igarashi, Y. Akazawa, K. Kawachi, S. Kagami, and M. Mochimaru, “An actuated physical puppet as an input device for controlling a digital manikin,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’11, pp. 637–646, 2011.
- [55] Z. Bhatti, I. A. Ismaili, S. Zardari, H. A. M. Malik, and M. Karbasi, “Procedural animation of 3D humanoid characters using trigonometric expressions,” *Bahria University Journal of Information & Communication Technologies*, vol. 9, no. 2, pp. 1–6, 2016.
- [56] Z. Bhatti, A. Shah, and F. Shahidi, “Procedural model of horse simulation,” in *Proceedings of the 12th ACM SIGGRAPH International Conference on*

Virtual-Reality Continuum and Its Applications in Industry, VRCAI '13, (New York, NY, USA), pp. 139–146, ACM, 2013.

- [57] V. Schwind, K. Leicht, S. Jäger, K. Wolf, and N. Henze, “Is there an uncanny valley of virtual animals? a quantitative and qualitative investigation,” *International Journal of Human-Computer Studies*, vol. 111, pp. 49–61, 2018.
- [58] A. A. A. Manaf, S. N. F. Ismail, and M. R. Arshad, “Perceived visual CGI familiarity towards uncanny valley theory in film,” *International Journal of Applied and Creative Arts*, vol. 2, no. 1, pp. 7–16, 2019.
- [59] M. N. Hamdan and A. Z. M. Ali, “User satisfaction of non-realistic three-dimensional talking-head animation courseware (3D-NR),” *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 5, no. 1, pp. 23–30, 2015.
- [60] E. Brochu, T. Brochu, and N. De Freitas, “A Bayesian interactive optimization approach to procedural animation design,” in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '10, (New York, NY, USA), pp. 103–112, ACM, 2010.
- [61] I. D. Horswill, “Lightweight procedural animation with believable physical interactions,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 1, no. 1, pp. 39–49, 2009.
- [62] F. Durupinar, U. Gdkbay, A. Aman, and N. I. Badler, “Psychological parameters for crowd simulation: From audiences to mobs,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 9, pp. 2145–2159, 2015.
- [63] 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.

- [64] S. Sonlu, U. Gdkbay, 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.
- [65] T. B. Moeslund, A. Hilton, and V. Krger, “A survey of advances in vision-based human motion capture and analysis,” *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.
- [66] I. A. Faisal, T. W. Purboyo, and A. S. R. Ansori, “A review of accelerometer sensor and gyroscope sensor in imu sensors on motion capture,” *Journal of Engineering and Applied Sciences*, vol. 15, no. 3, pp. 826–829, 2019.
- [67] Y. Desmarais, D. Mottet, P. Slangen, and P. Montesinos, “A review of 3D human pose estimation algorithms for markerless motion capture,” *Computer Vision and Image Understanding*, vol. 212, 2021. Article no. 103275, 19 pages.
- [68] S. Sharma, S. Verma, M. Kumar, and L. Sharma, “Use of motion capture in 3D animation: motion capture systems, challenges, and recent trends,” in *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*, COMITCon ’19, (Piscataway, NJ, USA), pp. 289–294, IEEE, 2019.
- [69] W. Qin, R. Tao, L. Sun, and K. Dong, “Muscle-driven virtual human motion generation approach based on deep reinforcement learning,” *Computer Animation and Virtual Worlds*, vol. 33, no. 3-4, 2022. Article no. e2092, 15 pages.
- [70] A. S. Anand, G. Zhao, H. Roth, and A. Seyfarth, “A deep reinforcement learning based approach towards generating human walking behavior with a neuromuscular model,” in *Proceedings of the IEEE-RAS 19th International Conference on Humanoid Robots*, Humanoids ’19, pp. 537–543, IEEE, 2019.
- [71] P. Xu and I. Karamouzas, “A GAN-like approach for physics-based imitation learning and interactive character control,” *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, vol. 4, Sept. 2021. Article no. 44, 22 pages.

- [72] J. Lee and H. Joo, “Locomotion-action-manipulation: Synthesizing human-scene interactions in complex 3D environments,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV ’23*, pp. 9663–9674, 2023.
- [73] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [74] L. Xu, Z. Song, D. Wang, J. Su, Z. Fang, C. Ding, W. Gan, Y. Yan, X. Jin, X. Yang, *et al.*, “ActFormer: A GAN-based transformer towards general action-conditioned 3D human motion generation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV ’23*, pp. 2228–2238, 2023.
- [75] 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.
- [76] Y. Liu, Q. Cao, Y. Wen, H. Jiang, and C. Ding, “Towards variable and coordinated holistic co-speech motion generation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR ’24*, pp. 1566–1576, 2024.
- [77] J. Ling, X. Tan, L. Chen, R. Li, Y. Zhang, S. Zhao, and L. Song, “Stableface: Analyzing and improving motion stability for talking face generation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 6, pp. 1232–1247, 2023.
- [78] B. H. Le, X. Ma, and Z. Deng, “Live speech driven head-and-eye motion generators,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 11, pp. 1902–1914, 2012.
- [79] Z. Ren, Z. Pan, X. Zhou, and L. Kang, “Diffusion motion: Generate text-guided 3D human motion by diffusion model,” in *Proceedings of the*

- IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '23*, pp. 1–5, IEEE, 2023.
- [80] Y. Yuan, J. Song, U. Iqbal, A. Vahdat, and J. Kautz, “PhysDiff: Physics-guided human motion diffusion model,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV '23*, pp. 16010–16021, 2023.
 - [81] K. Karunratanakul, K. Preechakul, S. Suwajanakorn, and S. Tang, “Guided motion diffusion for controllable human motion synthesis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV '23*, pp. 2151–2162, 2023.
 - [82] A. Amballa, G. Akkinapalli, and V. Muralikrishnan, “LS-GAN: Human motion synthesis with latent-space GANs,” in *Proceedings of the Winter Conference on Applications of Computer Vision*, pp. 326–335, 2025.
 - [83] Y. Wu, Z. Wu, and C. Ji, “Transformer-based partner dance motion generation,” *Engineering Applications of Artificial Intelligence*, vol. 139, Part B, 2025. Article no. 109610, 14 pages.
 - [84] J. Tanke, T. Shibuya, K. Uchida, K. Saito, and Y. Mitsufuji, “Dyadic Mamba: Long-term dyadic human motion synthesis,” in *Proceedings of the Second Workshop on Human Motion Generation, Co-located with the IEEE/CVF Conference on Computer Vision and Pattern Recognition, HuMoGen 2025*, pp. 2868–2877, 2025.
 - [85] G. Tevet, B. Gordon, A. Hertz, A. H. Bermano, and D. Cohen-Or, “Motion-CLIP: Exposing human motion generation to CLIP space,” in *Computer Vision – ECCV 2022* (S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds.), (Cham, Switzerland), pp. 358–374, Springer Nature, 2022.
 - [86] L. Pinheiro Cinelli, M. Araújo Marins, E. A. Barros da Silva, and S. Lima Netto, “Variational autoencoder,” in *Variational Methods for Machine Learning with Applications to Deep Networks*, pp. 111–149, Springer, 2021.

- [87] M. Petrovich, M. J. Black, and G. Varol, “Action-conditioned 3D human motion synthesis with Transformer VAE,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, CVPR ’21, pp. 10985–10995, 2021.
- [88] Z. Luo, S. A. Golestaneh, and K. M. Kitani, “3d human motion estimation via motion compression and refinement,” in *Computer Vision - ACCV 2020: 15th Asian Conference on Computer Vision, Kyoto, Japan, November 30 - December 4, 2020, Revised Selected Papers, Part V*, (Berlin, Heidelberg), pp. 324–340, Springer-Verlag, 2020.
- [89] X. Yan, A. Rastogi, R. Villegas, K. Sunkavalli, E. Shechtman, S. Hadap, E. Yumer, and H. Lee, “MT-VAE: Learning motion transformations to generate multimodal human dynamics,” in *Computer Vision - ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part V*, (Berlin, Heidelberg), pp. 276–293, Springer-Verlag, 2018.
- [90] M. Diomataris, N. Athanasiou, O. Taheri, X. Wang, O. Hilliges, and M. J. Black, “WANDR: Intention-guided human motion generation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR ’24, pp. 927–936, 2024.
- [91] E. Groff, “Laban movement analysis: Charting the ineffable domain of human movement,” *Journal of Physical Education, Recreation & Dance*, vol. 66, no. 2, pp. 27–30, 1995.
- [92] A. Aristidou, E. Stavrakis, P. Charalambous, Y. Chrysanthou, and S. L. Himona, “Folk dance evaluation using Laban movement analysis,” *Journal on Computing and Cultural Heritage*, vol. 8, no. 4, 2015. Article no. 20, 19 pages.
- [93] S. Chin, C. Y. Lee, and J. Lee, “An automatic method for motion capture-based exaggeration of facial expressions with personality types,” *Virtual Reality*, vol. 17, no. 3, pp. 219–237, 2013.

- [94] M. Koppensteiner, “Motion cues that make an impression: Predicting perceived personality by minimal motion information,” *Journal of Experimental Social Psychology*, vol. 49, no. 6, pp. 1137–1143, 2013.
- [95] M. Koppensteiner, “Perceiving personality in simple motion cues,” *Journal of Research in Personality*, vol. 45, no. 4, pp. 358–363, 2011.
- [96] R. E. Riggio, R. Lippa, and C. Salinas, “The display of personality in expressive movement,” *Journal of Research in Personality*, vol. 24, no. 1, pp. 16–31, 1990.
- [97] J. Hyde, E. J. Carter, S. Kiesler, and J. K. Hodgins, “Evaluating animated characters: facial motion magnitude influences personality perceptions,” *ACM Transactions on Applied Perception (TAP)*, vol. 13, no. 2, pp. 1–17, 2016.
- [98] M. Saberi, S. DiPaola, and U. Bernardet, “Expressing personality through non-verbal behaviour in real-time interaction,” *Frontiers in Psychology*, vol. 12, 2021. Article no. 660895, 19 pages.
- [99] W.-P. Su, B. Pham, and A. Wardhani, “Personality and emotion-based high-level control of affective story characters,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 2, pp. 281–293, 2007.
- [100] A. Shirvani, S. G. Ware, and L. J. Baker, “Personality and emotion in strong-story narrative planning,” *IEEE Transactions on Games*, vol. 15, no. 4, pp. 669–682, 2022.
- [101] F. Mairesse and M. Walker, “PERSONAGE: Personality generation for dialogue,” in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (A. Zaenen and A. van den Bosch, eds.), (Prague, Czech Republic), pp. 496–503, Association for Computational Linguistics, June 2007.
- [102] G. Jiang, M. Xu, S.-C. Zhu, W. Han, C. Zhang, and Y. Zhu, “Evaluating and inducing personality in pre-trained language models,” in *Advances in Neural Information Processing Systems*, vol. 36 of *NIPS ’23*, (Red Hook, NY, USA), pp. 10622–10643, Curran Associates, Inc., 2023.

- [103] M. P. Aylett, A. Vinciarelli, and M. Wester, “Speech synthesis for the generation of artificial personality,” *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 361–372, 2017.
- [104] Y. Lang, W. Liang, Y. Wang, and L.-F. Yu, “3d face synthesis driven by personality impression,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 1707–1714, 2019.
- [105] K. Legde and D. W. Cunningham, “Evaluating the effect of clothing and environment on the perceived personality of virtual avatars,” in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, IVA ’19, pp. 206–208, 2019.
- [106] D. Vogt, S. Grehl, E. Berger, H. Ben Amor, and B. Jung, “A data-driven method for real-time character animation in human-agent interaction,” in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, IVA ’14, pp. 463–476, Springer, 2014.
- [107] D. Delgado-Gómez, A. E. Masó-Besga, D. Aguado, V. J. Rubio, A. Sujar, and S. Bayona, “Automatic personality assessment through movement analysis,” *Sensors*, vol. 22, no. 10, 2022. Article no. 3949, 11 pages.
- [108] L. Cabrera-Quiros, E. Gedik, and H. Hung, “Multimodal self-assessed personality estimation during crowded mingle scenarios using wearables devices and cameras,” *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 46–59, 2019.
- [109] K. Ijuin and K. Jokinen, “Exploring gaze behaviour and perceived personality traits,” in *Social Computing and Social Media. Design, Ethics, User Behavior, and Social Network Analysis: 12th International Conference, SCSM 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19-24, 2020, Proceedings, Part I*, (Berlin, Heidelberg), pp. 504–512, Springer-Verlag, 2020.
- [110] J. C. J. Junior, Y. Güçlütürk, M. Pérez, U. Güçlü, C. Andujar, X. Baró, H. J. Escalante, I. Guyon, M. A. Van Gerven, R. Van Lier, *et al.*, “First

- impressions: A survey on vision-based apparent personality trait analysis,” *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 75–95, 2019.
- [111] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “BlazePose: On-device real-time body pose tracking,” 2020. Available at <https://arxiv.org/abs/2006.10204>, Accessed: 2025/08/11.
- [112] J. A. M. Correa, M. K. Abadi, N. Sebe, and I. Patras, “AMIGOS: A dataset for affect, personality and mood research on individuals and groups,” *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2021.
- [113] 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.
- [114] S. Ghorbani, K. Mahdavian, A. Thaler, K. Kording, D. J. Cook, G. Blohm, and N. F. Troje, “MoVi: A large multi-purpose human motion and video dataset,” *PloS ONE*, vol. 16, no. 6, 2021. Article no. e0253157, 15 pages.
- [115] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, “Berkeley MHAD: A comprehensive multimodal human action database,” in *Proceedings of the IEEE Workshop on Applications of Computer Vision, WACV '13*, (Piscataway, NJ, USA), pp. 53–60, IEEE, 2013.
- [116] C. Chen, R. Jafari, and N. Kehtarnavaz, “UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor,” in *Proceedings of the IEEE International Conference on Image Processing, ICIP '15*, (Piscataway, NJ, USA), pp. 168–172, IEEE, 2015.
- [117] C. van Gemeren, R. Poppe, and R. C. Veltkamp, “Spatio-temporal detection of fine-grained dyadic human interactions,” in *Human Behavior Understanding* (M. Chetouani, J. Cohn, and A. A. Salah, eds.), HBU '16, (Cham, Switzerland), pp. 116–133, Springer International Publishing, 2016.

- [118] K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, and D. Samaras, “Two-person interaction detection using body-pose features and multiple instance learning,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, CVPRW ’12, (Piscataway, NJ, USA), pp. 28–35, IEEE, 2012.
- [119] J. Liu, A. Shahroudy, M. Perez, G. Wang, L.-Y. Duan, and A. C. Kot, “NTU RGB+D 120: A large-scale benchmark for 3D human activity understanding,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2684–2701, 2019.
- [120] W.-R. Ko, M. Jang, J. Lee, and J. Kim, “AIR-Act2Act: Human–human interaction dataset for teaching non-verbal social behaviors to robots,” *The International Journal of Robotics Research*, vol. 40, no. 4-5, pp. 691–697, 2021.
- [121] G. Lee, Z. Deng, S. Ma, T. Shiratori, S. S. Srinivasa, and Y. Sheikh, “Talking with hands 16.2 m: A large-scale dataset of synchronized body-finger motion and audio for conversational motion analysis and synthesis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, CVPR ’19, (Piscataway, NJ, USA), pp. 763–772, IEEE, 2019.
- [122] 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.
- [123] L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. Lau, *et al.*, “Conversational agents in healthcare: a systematic review,” *Journal of the American Medical Informatics Association*, vol. 25, no. 9, pp. 1248–1258, 2018.
- [124] H. Jiang, X. Zhang, X. Cao, C. Breazeal, D. Roy, and J. Kabbara, “Person-aLLM: Investigating the ability of large language models to express personality traits,” in *Findings of the Association for Computational Linguistics: NAACL 2024* (K. Duh, H. Gomez, and S. Bethard, eds.), (Mexico City, Mexico), pp. 3605–3627, Association for Computational Linguistics, June 2024.

- [125] Meta Horizon, “Oculus Lipsync Unity,” 2025. Available at <https://developers.meta.com/horizon/downloads/package/oculus-lipsync-unity/>, Accessed: 2025/08/11.
- [126] IBM Corporation, “IBM Watson API,” 2015. Available at <https://www.ibm.com/watson>, Accessed: 2025/08/11.
- [127] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, “A very brief measure of the big-five personality domains,” *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, 2003.
- [128] J. Musek, “A general factor of personality: Evidence for the big one in the five-factor model,” *Journal of Research in Personality*, vol. 41, no. 6, pp. 1213–1233, 2007.
- [129] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, *et al.*, “A survey on evaluation of large language models,” *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 3, 2024. Article no. 39, 45 pages.
- [130] B. Wang, G. Li, and Y. Li, “Enabling conversational interaction with mobile UI using large language models,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI ’23, (New York, NY, USA), ACM, 2023. Article no. 432, 17 pages.
- [131] G. Serapio-García, M. Safdari, C. Crepy, L. Sun, S. Fitz, M. Abdulhai, A. Faust, and M. Matarić, “Personality traits in large language models,” 2025. Available at <https://arxiv.org/abs/2307.00184>, Accessed: 2025/08/11.
- [132] K. Pan and Y. Zeng, “Do LLMs possess a personality? making the MBTI test an amazing evaluation for large language models,” 2023. Available at <https://arxiv.org/abs/2307.16180>, Accessed: 2025/08/11.
- [133] A. Dregger, M. Seifermann, and A. Oberweis, “Language cues for expressing artificial personality: A systematic literature review for conversational agents,” in *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, CUI ’24, 2024. Article no. 22, 17 pages.

- [134] Reallusion, “Character Creator: 3D character design software,” 2025. Available at <https://www.reallusion.com/character-creator/>, Accessed: 2025/08/11.
- [135] C. J. Soto and O. P. John, “Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS,” *Journal of Research in Personality*, vol. 68, pp. 69–81, 2017.
- [136] R. H. Kay and L. Knaack, “Assessing learning, quality and engagement in learning objects: the Learning Object Evaluation Scale for Students (LOES-S),” *Educational Technology Research and Development*, vol. 57, pp. 147–168, 2009.
- [137] H. Zhou and T. Zhang, “Body language in business negotiation,” *International Journal of Business and Management*, vol. 3, no. 2, pp. 90–96, 2008.
- [138] M. Romeo, D. Hernández García, T. Han, A. Cangelosi, and K. Jokinen, “Predicting apparent personality from body language: benchmarking deep learning architectures for adaptive social human–robot interaction,” *Advanced Robotics*, vol. 35, no. 19, pp. 1167–1179, 2021.
- [139] H. Winther, “Body contact and body language: Moments of personal development and social and cultural learning processes in movement psychology and education,” *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, vol. 9, no. 2, 2008. Article no. 63, 23 pages.
- [140] S. Kita, “Cross-cultural variation of speech-accompanying gesture: A review,” *Language and Cognitive Processes*, vol. 24, no. 2, pp. 145–167, 2009.
- [141] A. Lücking and J. Ginzburg, *Saying and shaking‘No’*. Frankfurt, Germany: Universitätsbibliothek Johann Christian Senckenberg, 2021.
- [142] C. Zlogar and K. Davidson, “Effects of linguistic context on the acceptability of co-speech gestures,” *Glossa: A Journal of General Linguistics*, vol. 3, no. 1, 2018. Article no. 73, 28 pages.

- [143] R. E. Ready, L. A. Clark, D. Watson, and K. Westerhouse, “Self-and peer-reported personality: Agreement, trait ratatability, and the “self-based heuristic”,” *Journal of Research in Personality*, vol. 34, no. 2, pp. 208–224, 2000.
- [144] C. Chen, Y. Zhuang, F. Nie, Y. Yang, F. Wu, and J. Xiao, “Learning a 3D human pose distance metric from geometric pose descriptor,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 11, pp. 1676–1689, 2010.
- [145] W. Zhu, X. Ma, D. Ro, H. Ci, J. Zhang, J. Shi, F. Gao, Q. Tian, and Y. Wang, “Human motion generation: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 4, pp. 2430–2449, 2023.
- [146] Khronos Group, “WebGL: Low-level 3D graphics API based on OpenGL ES,” 2025. Available at <https://www.khronos.org/webgl/>, Accessed: 2025/08/11.
- [147] R. Cabello and three.js Developers, “Three.js - JavaScript 3D Library,” 2025. Available at <https://threejs.org/>, Accessed: 2025/08/11.
- [148] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS ’17*, (Red Hook, NY, USA), pp. 6000–6010, Curran Associates, Inc., 2017.
- [149] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, “On the variance of the adaptive learning rate and beyond,” *CoRR*, vol. abs/1908.03265, 2019. Available at <http://arxiv.org/abs/1908.03265>, Accessed: 2025/08/11.
- [150] A. L. Jones, R. S. Kramer, and R. Ward, “Signals of personality and health: the contributions of facial shape, skin texture, and viewing angle,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 38, no. 6, pp. 1353–1361, 2012.