Check for updates

Personality Expression Using Co-Speech Gesture

SİNAN SONLU, HALİL ÖZGÜR DEMİR, and UĞUR GÜDÜKBAY, Bilkent University, Ankara, Turkey

We express our personality through verbal and nonverbal behavior. While verbal cues are mostly related to the semantics of what we say, nonverbal cues include our posture, gestures, and facial expressions. Appropriate expression of these behavioral elements improves conversational virtual agents' communication capabilities and realism. Although previous studies focus on co-speech gesture generation, they do not consider the personality aspect of the synthesized animations. We show that automatically generated co-speech gestures naturally express personality traits, and heuristics-based adjustments for such animations can further improve personality expression. To this end, we present a framework for enhancing co-speech gestures with the different personalities of the Five-Factor model. Our experiments suggest that users perceive increased realism and improved personality expression when combining heuristics-based motion adjustments with co-speech gestures.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; • **Computing methodologies → Animation**;

Additional Key Words and Phrases: OCEAN personality, human animation, conversational agent, co-speech gesture

ACM Reference format:
Sinan Sonlu, Halil Özgür Demir, and Uğur Güdükbay. 2024. Personality Expression Using Co-Speech Gesture. *ACM Trans. Appl. Percept.* 22, 2, Article 7 (November 2024), 20 pages.
https://doi.org/10.1145/3694905

1 Introduction

Personality gives soul to our otherwise lifeless conversations. Whether consciously or not, we put a part of ourselves into our words, gestures, and facial expressions [39]. The clothes we wear [53] and the length of our hair [9, 56] all give information about who we are. We expect the same from our virtual characters, which appear dull when we leave out personality and emotions. Virtual characters require a close inspection of the different communication channels for successful and believable performances. A thoughtful design of body motion is essential for expressing the desired personality in conversational virtual agents [18, 63].

Avoiding eye contact may signal insecurity; sloppy hand movements could mean irresponsibility; a rising posture would reflect optimism. When we speak, our motion accompanies our words, and the style of these movements reflects the inner self. We examine the power of co-speech gestures in the personality expression of virtual characters, utilizing a recent co-speech gesture synthesis model [25] and heuristics-based animation

Sinan Sonlu and Halil Özgür Demir contributed equally to this research.

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

Authors' Contact Information: Sinan Sonlu, Bilkent University, Ankara, Turkey; e-mail: sinan.sonlu@bilkent.edu.tr; Halil Özgür Demir, Bilkent University, Ankara, Turkey; e-mail: hozgurde@gmail.com; Uğur Güdükbay (corresponding author), Bilkent University, Ankara, Turkey; e-mail: gudukbay@cs.bilkent.edu.tr.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 1544-3965/2024/11-ART7
https://doi.org/10.1145/3694905

transformations to enhance personality expression [18, 65]. We report the generated animations' performance in conveying the desired personality using different combinations of co-speech gestures and heuristics-based enhancements.

Although personality expression happens through multiple channels [19], and the different communication channels do not always have to agree, various traits often co-occur. For example, the face and body of a virtual character may reflect opposing personality traits, yet in real life, these channels are usually consistent. Our speech, gesture, and facial expressions share common characteristics; they tell the same story to different receptors [38]. As a result, we expect a co-speech gesture model trained on real-life data to capture such intricate connections; words common to a specific personality would encourage gestures that signal the same type of personality. Similarly, vocal features indicate different personalities [60], and voice qualities can drive personality-specific motion. Moreover, data-driven systems can automatically extract control parameters to represent different motion characteristics using measurable properties such as average gesture speed, height, spatial extent, and lateral symmetry [4]. Co-speech gestures of different styles generated with the same speech input can help portray different personality traits. Our experiments test if these naturally occurring cues are visible to the average user and whether existing methods can improve the personality expression in such automatically generated animations.

We conduct a user study comparing different versions of the same conversation about personality expression and realism. We utilize the personality-specific dialogues of an existing system [65] and generate co-speech gestures to accompany these dialogues using **Zero-Shot Example-Based Gesture Generation from Speech (ZeroEGGS)** [25]. The different versions include neutral, happy, and sad co-speech gestures. We also utilize personality-based motion adjustments on the neutral co-speech samples to enhance personality expression as a different version, and the final version of the conversation uses the same repeated talking animation with personality-based modifications. Our personality-based motion adjustments utilize **Laban Movement Analysis (LMA)** Efforts as a theoretical basis. We replicate the animation using **Inverse Kinematics (IK)** and shift hand end effectors along different directions to represent LMA Space and Weight Efforts. IK interpolation speed reflects the changes in LMA Time, and random noise is added to joints to show LMA Flow. In contrast to the previous work that requires a preprocessing on the whole animation to apply personality-based modifications [65], we use a simplified algorithm that can run online.

We compare the different versions of the conversations that focus on expressing specific personality traits. Our results suggest that combining co-speech gestures with heuristics-based personality modifications improves the portrayal of high traits. Co-speech gestures with a neutral style have a good average performance when we target expressing traits in isolation. We observe that personality-based motion adjustments mostly influence the perceived personality dimensions together, helping achieve a high distinction for the opposite traits at the cost of increasing correlation. Utilizing happy or sad style co-speech gestures helps express the desired traits naturally for specific factors.

Our contribution includes a system for enhancing co-speech gestures for better expression of the personality types of the Five-Factor model. We analyze the success of the resulting system via a detailed user study, exploring the benefits of utilizing co-speech gestures. Our framework and results are available in our GitHub repository[1] for further studies in personality expression using co-speech gestures. Please check our supplementary video for a visual summary of our approach and results.

2 Related Work

To our knowledge, current data-driven animation generation systems do not support personality-specific motion synthesis. Various studies utilize training data labels to generate animations expressing different styles. These styles often include emotional categories such as happy, sad, and angry [59] or generic categories such as childlike,

[1] https://github.com/hozgurde/Interactive-Agents

depressed, and old [62]. Studies also utilize videos as input style [1], where the network transforms the input animation to express the target style; in this case, the input animation and video should include similar motion content; for example, an energetic walk can be transferred into a depressed walk. Style can also be inferred automatically using measurable parameters for controlling the output characteristics of the generated motion [4].

The literature on emotion understanding can inspire studies in personality analysis and expression. For example, unintentional behaviors driven by inner feelings, like touching the head or playing with hair, can reveal the emotional states behind these micro gestures [45]. Automatically detecting such subtle pose details is possible using multi-scale 3D-shift graph convolutional networks that enable interactions of joints within a spatial-temporal volume for global feature extraction [61]. In addition to the pose and visual features, data augmentation [24] and utilizing textual descriptions of the gestures [42] help improve the analysis performance. Datasets such as SMG [14], which connect spontaneous body gestures to emotional stress states, could provide interesting implications for neuroticism analysis. While emotional gestures emerge spontaneously and vary wildly, a temporal analysis of such features can reveal connections to personality; to this end, emotional analysis focusing on long sequences [41] is more applicable to personality analysis.

2.1 Data-Driven Animation

Early examples of data-driven motion generation use statistical models to generalize the captured motion data for new tasks [50]. Generating novel motion samples usually requires an input such as the expected motion trajectories [34], targeted task [57], or an input animation to transfer its style [1]. Recent diffusion models utilize descriptive text input for motion generation [66, 70]. Data annotations form the limitation of most data-driven approaches. Such systems can generate motion in predefined categories [59] or according to hand-crafted qualities [22]; however, introducing new categories to these systems requires relabeling and retraining.

Various studies utilize input speech to generate accompanying animation. In this case, a reference style can be used as an input to generate accompanying gestures, allowing the final motion to be associated with this reference style [25]. This study explores the extent of personality expression when using co-speech gestures with neutral, happy, and sad styles. A different approach is to create a style matrix while training the model with both speech and style data to embed style into speech during inference [3]. It is also possible to extract specific parameters from speech data, creating gesture slots based on the length of the speech to select the best-fitting gestures from a database for these slots [23]. These gestures can be converted into animation using interpolation. Speech input can also drive facial animations [13], controlling the mouth shape and expression.

Data-driven co-speech gesture synthesis can utilize different input types, including audio, text, and pose data, as summarized by Nyatsanga et al. [54]. Generated gestures can follow a rhythm to increase the sense of realism [5] or may expect a target position to generate context-aware pointing gestures using motion imitation [15]. Considering the semantic meaning of the driving speech can help achieve more appropriate gesturing [44], and utilizing motion prompts with the accompanying speech enables more controlled characteristics for the generated animations [28]. Although many models require samples to train generative systems that can output stylized motion that resembles the target speaker, DiffGAN enables low-resource adaptation for personalized co-speech gesture generation [2]. Also, models can input an interacting partner's movements to generate gestures in dyadic interaction scenarios [67].

2.2 Parametric Animation

While early animation is mostly hand-crafted, the task's difficulty necessitates automation. Utilizing IK can help reduce the parameters involved [51]. In this case, the animator controls the end effectors' positions, and the intermediate joints' rotations are automatically determined. Procedural animation involves defining the motion using parametric equations [33] and enables the generation of samples of high variance, expressing different styles. Although studies represent cyclic motions such as walking and running mathematically [10], expressing generic animations procedurally is challenging and may result in unnatural motions.

A different approach is to adjust an existing animation parametrically to represent different styles [32], usually to express personality traits [18, 65] or emotions [16, 68]. Expert-driven perceptual features utilizing Gaussian radial basis functions to joint trajectories enable altering the emotional style of neutral motion samples [22]. Referring to animation and motion experts is common in affective motion synthesis; for example, PERFORM [18] utilizes LMA experts to develop motion parameters to represent LMA qualities. Animation adjustment techniques include speed scaling, applying additive rotations to change posture, and modifying joint movement trajectories. These adjustments are generally based on a high-level understanding of motion; LMA is one such framework, offering interpretable parameters to define and adjust human motion [37]. LMA Efforts of Space, Time, Flow, and Weight can be used for analyzing and altering existing motion. A limitation of this approach is that the output heavily depends on the input animation. The heuristics-based modifications usually have assumptions on the input, such as requiring the input animation to have a neutral personality [65]. Previous work establishes connections between LMA Efforts and different personalities of the Five-Factor Model [18], which we utilize in this study to enhance personality expression.

2.3 Personality Expression

Although there are attempts at personality-labeled animation datasets [64], most of the existing motion datasets do not include personality labels [25, 26, 31], and those with personality annotations lack full-body motion [12, 17], which is crucial for the gesture. Whether intentional or not, all behaviors expose psychological cues [20, 49]; consequently, we expect any human motion dataset that utilizes real-life motion capture to contain a personality layer. Moreover, this personality layer can be inferred from existing labels. For example, we can infer personality labels from emotion annotations using the relationship between emotions and personality [35]. Similarly, speech and gesture are related [11], and we can exploit the links between dialogue-personality and voice-personality to generate speech-driven personality motion. Although we expect co-speech gestures generated from personality-enriched speech to express personality, further adjustments to these animations can improve the expression of the desired personality to a greater extent.

Many methods exist for expressing personality in digital characters. Dialogue text by itself can express personality [43]; while the most popular approach for personality expression is through facial animation [6, 30] and full-body movement [18, 27, 63], factors such as body shape [29] are also influential. Motion and body shape together can influence the perceived sex of virtual characters [48]. Employing different hand motions can alter the apparent personality [69]. With all these different approaches, preserving the relationship between shape and motion is essential, as animation inconsistencies harm the perceived realism dramatically [36]. Realistic rendering of the virtual agents [71] and appropriate gesturing can help increase immersion.

In this study, we employ a multi-modal approach to personality expression utilizing speech-driven gesture animations and heuristics-based modifications. In our conversation animations, we utilize existing dialogue and accompanying speech respecting the relationship between vocal features and personality [58], which is designed for each extreme personality of the Five-Factor model in an earlier study [65]. We compare the personality expression, realism, and motion appropriateness of five animation alternatives that accompany the same speech, with each animation utilizing a different approach to portray the target personality.

3 Framework

We propose a framework, summarized in Figure 1, that combines co-speech gestures with heuristics-based motion adjustments to express the desired personality traits in conversational agents. Our framework utilizes co-speech gestures generated by ZeroEGGS, which outputs a **Biovision Hierarchy (BVH)** animation file accompanying the input speech. We use a custom BVH importer to adapt the generated animation to our **three-dimensional**

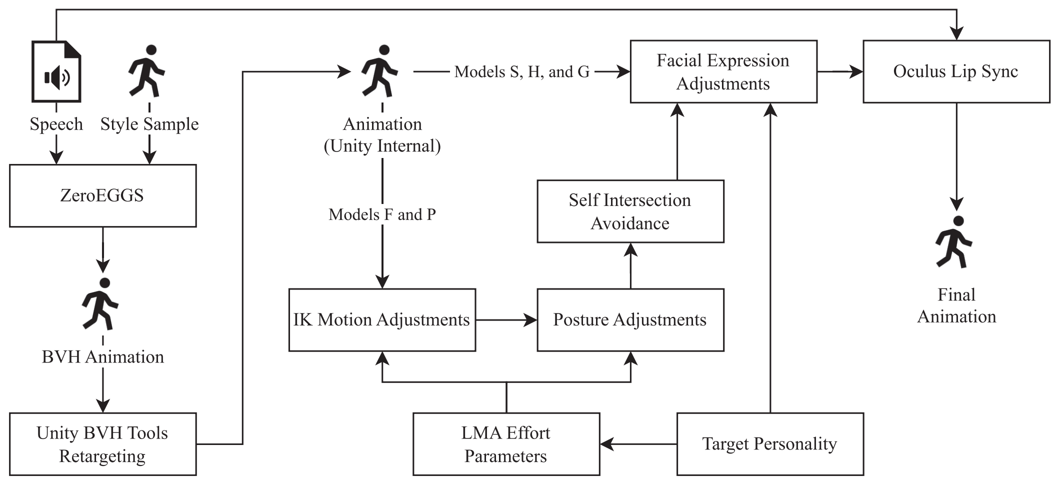


Fig. 1. The proposed framework for combining co-speech gestures with heuristics-based motion adjustments. We utilize personality-specific dialogue speech for each model to synthesize co-speech gesture animation using ZeroEGGS. Models F, P, and G use the neutral style sample, while Model H uses the happy and Model S uses the sad style samples. Model P uses the same BVH animation regardless of the agent's speech (we generate one gesture animation to be used in all samples of Model P); all other models use the BVH animation specific to the current speech of the agent. Using BVH Tools and animation retargeting ensures samples are displayed correctly on the agent. All models utilize personality-specific facial expression adjustments and lip sync. Models F and P utilize IK-based motion adjustments and additive rotations to modify the general posture based on LMA Effort parameters. Since Models F and P change the animation, we apply collision checks on the agent's limbs and shift any colliding parts to the closest point on limb surfaces.

(3D) model. We map the joints of the animation to the 3D model's joints and retarget the bone rotations while respecting the rest of the armatures. This enables viewing unmodified ZeroEGGS animations in Unity using our agent model. Three models we compare in our experiment utilize the ZeroEGGS animations with no heuristics-based motion adjustments. We assume co-speech gestures naturally possess personality cues and observe if these are apparent to the average viewer. Two of the models utilize heuristics-based motion adjustments that are added on top of the current animation of the agent based on the target personality. We repurpose techniques from our previous work [65]. In particular, we modify the gestures using IK-based retargeting, additive rotations for changing posture, general speed changes, and noise addition. All modifications utilize LMA Efforts as a basis.

3.1 Co-Speech Gesture Generation

ZeroEGGS can generate 3D co-speech animations using input speech and a sample style. While input speech drives the overall gesturing of the generated animation, the input style controls the motion qualities of the output without explicit mapping. Using representative samples from the ZeroEGGS dataset, we generate animations with neutral, happy, and sad styles to compare their performances in conveying personality traits.

We use ZeroEGGS to generate three sets of animations (neutral, happy, and sad styles) using the speech files of the passport scenario [65]. Since the agent's speech already includes custom dialogue and vocal features to express each personality of the Five-Factor Model, we expect the generated co-speech animations to accommodate personality traits. Unity uses Filmbox format to represent animations internally; however, since ZeroEGGS outputs BVH files, we use BVH Tools for Unity [21] to load the animations with a custom bone mapping. Additionally, the resting poses of the generated animations and our 3D model do not match (see Figure 2); consequently, we retarget rotations of the animation respecting the difference in the rest poses of the armatures; in particular, we subtract the difference in their rotations in each frame of animation.

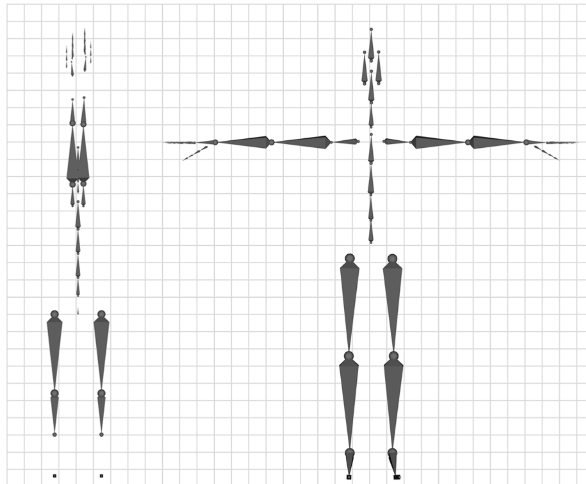


Fig. 2. Rest pose of ZeroEGGS output (left) and our model (right).

3.2 Personality Adjustments

Two of the models in our experiment utilize heuristics-based motion adjustments to enhance the perception of the target personality traits through expressive gesturing. With some changes, we adapt an existing solution [65] to our framework. First, the existing solution uses an initial pass on the original animation to determine attraction weights to move hands toward a point in space so as not to change the resulting motion drastically. However, we ultimately aim for an online solution to generate co-speech gestures in real time, preventing observing the whole animation before playback. Consequently, our IK-based gesture changes shift the hand end effectors in each frame by an amount determined by LMA Space and Weight. Indirect Space causes the hands to be shifted outwards, while Direct Space shifts the hands inwards; similarly, Heavy Weight shifts the hands downwards while Light Weight shifts upwards. We reproduce the current pose of the input animation by mimicking its hand positions with IK and then shift the end effectors toward the aforementioned direction to alter its observed LMA Space or Weight.

Second, the existing LMA Time modifications directly influence the animation speed of the agent since the utilized base animations do not include a meaningful connection between the gestures and dialogue. However, one main advantage of using data-driven co-speech gestures is that the agent's motion strongly connects to its speech. Therefore, we avoid breaking this connection by directly altering the animation speed. Instead, we alter the interpolation speed of the IK system, causing Sustained Time to follow the original animation slightly late as opposed to Sudden Time following the original animation strictly. This could reduce the impact of the changes due to LMA Time, whereas the correspondence between gestures and speech is preserved.

Third, the system depends on manually determined adjustment weights to avoid self-intersections. For example, attracting the hands inwards may cause the hands to intersect with the torso, which was prevented by lowering the attraction factor for specific animations in the original work. This study introduces body awareness for hand motion to support altering a wide range of co-speech animations; we attach colliders to the model's body, arms, and hands to achieve this. We check for collisions at each frame using the adjusted positions of the model's hands and arms. If there are self-intersections, the intersecting part of the limb is shifted to the closest point on the intersected body's surface, avoiding directly using the transformations on the arms intact.

The theoretical basis of the existing system depends on PERFORM [18], an earlier solution for adding personality to human motion utilizing LMA. PERFORM relies upon two LMA experts to establish the motion parameters that reflect different LMA features; then, personality implications of the various LMA parameters are analyzed and validated through perception studies. The existing system and the changes made for this work respect the findings of PERFORM while extending its application to co-speech animations.

The existing system utilizes facial expressions to support the target personality when the heuristics-based modifications are enabled; to neutralize any change in perception due to facial expressions, we use the same facial expressions in each model regardless of whether the motion adjustments are active. Additionally, we keep the same neutral finger movements in all models to focus on the differences in gestures. All models utilize mouth animations that follow the agent's speech using Oculus Lip Sync [47].

Certain aspects of the existing system are directly used in this study; these include additive per-joint noise due to LMA Flow, static rotations to adjust the general posture of the spine due to LMA Weight and blink speed, and facial expression decay rate related to neuroticism. We also utilize the same quantitative mapping between personality factors and LMA Effort parameters as

$$\begin{aligned} \text{LMA}_{Space} &= O + E - C - N \\ \text{LMA}_{Weight} &= O + E + A - N \\ \text{LMA}_{Flow} &= N + E + O - C \\ \text{LMA}_{Time} &= E + N - C. \end{aligned} \tag{1}$$

The equations determine direct and inversely proportional quantities since we focus on expressing one factor at a time. Positive LMA_{Space}, LMA_{Weight}, LMA_{Flow}, and LMA_{Time} correspond to Indirect Space, Light Weight, Free Flow, and Sudden Time, respectively. We summarize the agent's gesture changes due to heuristics-based motion adjustments in Figure 3.

4 Experiments

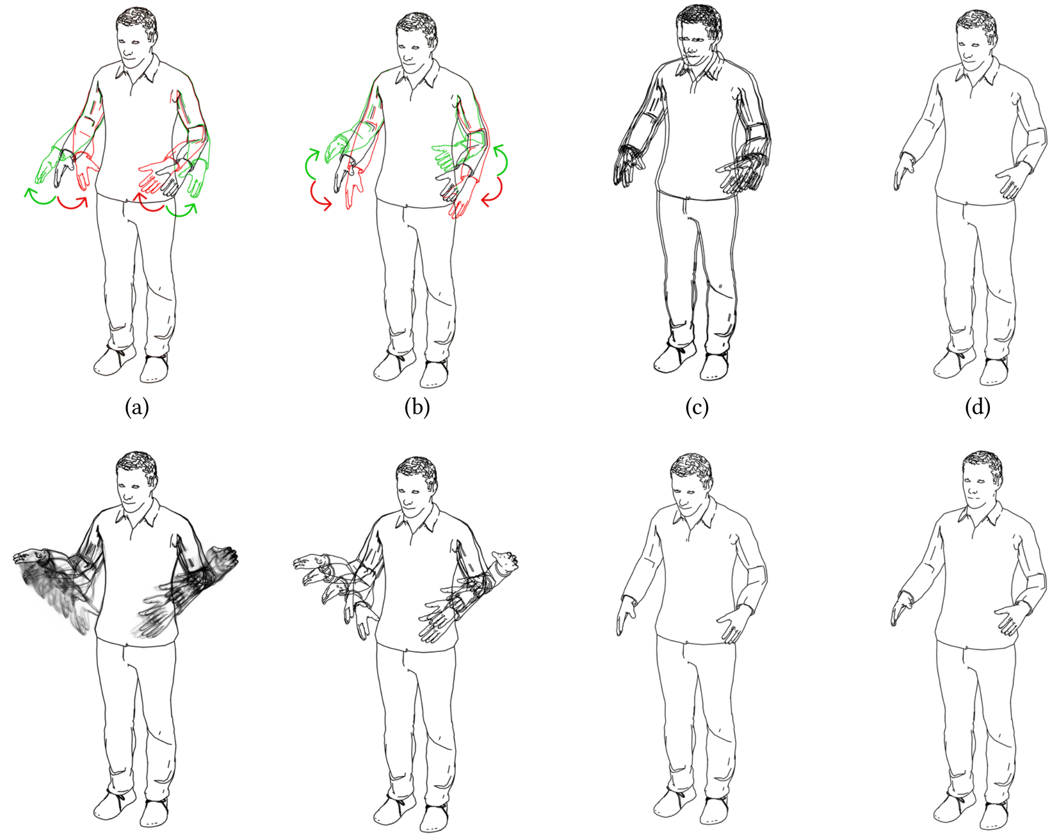
We conducted a user study[2] to evaluate the performance of our framework in representing different personalities of the Five-Factor Model. We measured apparent personality, human likeness, and appropriateness of one-minute videos of the conversational agent. The agent animations had five variations following different strategies to compare our approach to other strategies. We follow the experimental setup of an existing work [65] where the agent speaks with a passport officer using personality-specific dialogue, hand-crafted to represent each polarity of the Five-Factor Model. We use the following models to form different agent animation variations:

— *Model F* uses our full pipeline, combining co-speech gestures generated with the neutral style with real-time personality-based motion adjustments.
— *Model G* uses unmodified co-speech gestures generated with the neutral style only. The performance of this model indicates how well speech-driven gestures capture the individual's personality.
— *Model P* uses real-time personality-based motion adjustments without co-speech gestures. We use the same base animation for each sample, a neutral talking animation. This model is equivalent to the personality expression system of Sonlu et al. [65], where the same base motion is used for expressing each personality.
— *Model H* is similar to Model G but differs in the co-speech gesture generation style input. This model uses the happy style to generate the agent animations, resulting in more energetic and active movements.
— *Model S* is similar to Model G but differs in the co-speech gesture generation style input. This model uses the sad style to generate the agent animations, resulting in slow and passive movements.

4.1 Data Preparation

We utilize the dialogue (including the generated speech) and the setting (including the models for the scene and agent) of the Passport Scenario from Sonlu et al. [65], where they use hand-crafted dialogue for each agent

[2] Approved by the Ethical Committee of Bilkent University, Decision No: 2023_06_09_02.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3. The visual summary of the motion adjustments due to each LMA Effort. These adjustments alter agent animations to enhance the expression of different personality traits. LMA Space (a) shifts hand trajectories outwards for Indirect Space and inwards for Direct Space. LMA Weight (b) shifts hand trajectories upwards for Light Weight and downwards for Strong Weight. LMA Flow adds random noise to each joint's local rotation; Free Flow results in shaky motion (c), whereas Bound Flow does not add noise and results in a steady motion (d). LMA Time contributes to IK interpolation speed: Quick Time uses fast interpolation that results in speedy gestures (e); on the other hand, Sustained Time uses slow interpolation that causes the same gestures to appear unhurried (f). LMA Weight also contributes to the general posture of the body; Strong Weight causes the spine and neck to tilt downwards (g), while Light Weight causes these joints to tilt upwards (h). These modifications are applied to the agent's animation in real-time for Models F and P.

focusing on one extreme personality. The speech of the agents utilizes the text-to-speech API of IBM Watson [46]. To generate the agent animations for Models F, G, H, and S, we input these speech files to ZeroEGGS. Models F and G use the neutral style, Model H uses the happy style, and Model S uses the sad style. Model P repeats the same generic speech animation for all dialogue lines for each personality. Models F and P utilize real-time motion adjustments to enhance personality expression. We prepare ten videos for each model, each focusing on one personality type, expressing the corresponding trait highly or lowly. In each video, the upper body of the agent is shown using a static camera during a dialogue with the passport officer. We use these videos in our user study.

Fig. 4. Sample screenshot from the experiment showing five agent videos with the same target personality. The participants rate the agent in each video using the sliders below.

4.2 Experimental Setup

We use our online survey website to perform the user study. The participants are recruited from the crowd-sourcing service Prolific and are paid a fixed amount to complete the study. We use the demographics provided by Prolific and do not ask for any personal information from the participants.

The experiment involves 10 tasks, each including the 5 animation variations that aim to express the same personality. The tasks are randomly assigned to the participants; each participant completes all 10 tasks. In each task, the participants watched five randomly ordered videos in a simultaneous manner. The sound of the videos is muted except for one to ensure better audio quality. The videos use the vertical format, focusing on the agent, so the participant easily view them side by side. Participants can replay, pause, and resume the videos simultaneously and control the individual videos similarly.

We ask the participants to rate each video using the sliders below them following seven measurements. Five measurements use the personality traits from the **Ten-Item-Personality-Inventory (TIPI)** questionnaire, where opposite polarity traits appear on the opposite sides of the slider, similar to how PERFORM [18] uses TIPI. The othertwo 2 measurements focus on human likeness and motion appropriateness [40]. Each measurement type is explained to the participants in the introductory text before and after accepting participation. The sliders represent integer values between -100 and 100, and the slider's current value is shown beneath, together with changes in color and label size for a more interactive user experience. We mark the background of the sliders on a 7-point scale, as shown in Figure 4. The slider values are from a wide range to ease the comparison between the samples of similar appearance. For example, even if the participant considers each agent to have high extraversion, there will be room to rate them slightly differently.

Although each sample focuses on changes in one personality dimension, participants rate all five factors. Ultimately, we receive a score in the range $[-100,100]$ for each personality factor, human likeness, and motion appropriateness per sample. We normalize these measurements to the $[-1,1]$ range for the analysis. We expect the samples focusing on opposing personalities to have a significantly high difference for the related factor. For example, we expect extroverted and introverted samples to have a high difference in their mean extroversion scores.

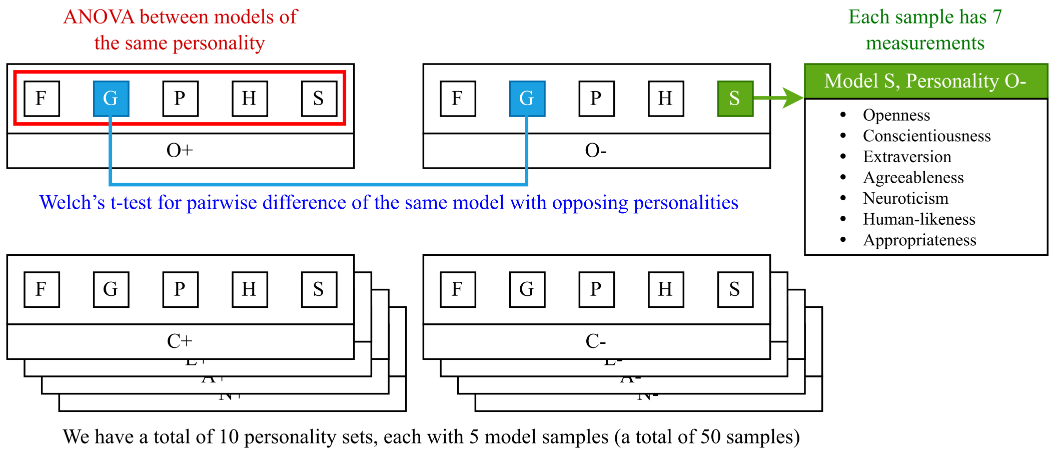


Fig. 5. Summary of our data analysis.

4.3 Data Analysis

We summarize our data analysis in Figure 5. We have 10 personality sets expressing one target personality from $\{O+, O-, C+, C-, E+, E-, A+, A-, N+, N-\}$. Each personality set has a sample from each model, marked with the model's name as $\{F, G, P, H, S\}$. For each sample, we collect personality measurements for openness (O), conscientiousness (O), extraversion (E), agreeableness (A), neuroticism (N), as well as measurements for human likeness (L), and motion appropriateness (M). We perform one-way ANOVA with Tukey Honestly Significant Difference to reveal significant differences between different models for each personality set and measurement type. The ANOVA results help reveal the most suitable model for expressing the traits of each personality type.

We can also determine the model performance by the distinction it achieves in representing the opposite personality traits. We perform Welch's t-test on the participants' ratings for each sample of opposing traits; we report the p-values showing the statistical significance of the differences per factor and model. Significant statistical differences in distinguishing the opposing traits for a personality factor signal a good performance for the corresponding model. As we make multiple comparisons, we adjust the p-values to correct Type 1 errors using False Discovery Rate control [8]. We perform principal component analysis to investigate the correlation between the different personality factors of the samples for each model and analyze the linear correlations between the measurements of each model.

5 Results

About 57 unique users (30 female, 27 male) participated in our study with a mean age of 29.36 ± 9.09. Participants are from different countries, with 27 being from European countries (Poland 9, Portugal 4, United Kingdom 2, Greece 2, Hungary 1, Italy 2, and other countries), 20 from Africa, 7 from America, and the rest from other continents. Each user rated all the samples in different personality sets that appeared randomly. The average time to complete the experiment was 31.26 ± 13.29 minutes.

5.1 Variance Analysis

ANOVA statistics for each personality set are given in Table 1. We use the model name as the independent variable and the different types of measurements as the dependent variables for calculating ANOVA statistics per measurement. Sets of high agreeableness ($A+$) and high conscientiousness ($C+$) have significant differences between models for almost all measurement types. On the other hand, the low openness ($O-$) and low conscientiousness ($C-$) sets have the least amount of measurements that have significant differences. The measurements with high

Table 1. ANOVA Statistics for Each Personality Set Where the Used Model is the Independent Variable and the Participants' Ratings for Each Type of Measurement (Meas.) is the Dependent Variable

High Polarity (+)						Low Polarity (-)					
Set	**Meas.**	**Between Groups**			**Within Groups SS**	**Set**	**Meas.**	**Between Groups**			**Within Groups SS**
		SS	**F**	η^2				**SS**	**F**	η^2	
O+	**O**	11.25	14.547	0.172	54.15	**O-**	**O**	3.31	3.540	0.048	65.42
	C	0.48	0.646	0.009	52.26		**C**	1.50	1.660	0.023	63.12
	E	18.82	21.859	0.238	60.28		**E**	5.18	6.387	0.084	56.73
	A	4.47	5.378	0.071	58.24		**A**	0.85	0.947	0.013	63.14
	N	0.88	0.975	0.014	63.20		**N**	0.65	0.717	0.010	63.49
	L	3.44	3.156	0.043	76.38		**L**	0.26	0.223	0.003	83.00
	M	2.81	2.856	0.039	68.83		**M**	0.43	0.453	0.006	66.40
C+	**O**	7.73	9.243	0.117	58.54	**C-**	**O**	6.30	8.085	0.104	54.53
	C	2.21	3.122	0.043	49.54		**C**	0.98	1.140	0.016	60.22
	E	11.24	10.883	0.135	72.29		**E**	8.12	9.856	0.123	57.63
	A	4.45	5.852	0.077	53.18		**A**	0.69	0.850	0.012	57.20
	N	2.15	2.575	0.035	58.44		**N**	2.05	1.768	0.025	81.18
	L	5.16	4.471	0.060	80.86		**L**	1.51	1.449	0.020	72.99
	M	2.05	2.078	0.029	69.22		**M**	0.22	0.233	0.003	64.90
E+	**O**	14.93	18.728	0.211	55.81	**E-**	**O**	7.50	7.539	0.097	69.67
	C	0.32	0.357	0.005	63.29		**C**	2.34	2.402	0.033	68.23
	E	15.58	16.427	0.190	66.40		**E**	10.00	9.149	0.116	76.53
	A	4.22	4.824	0.064	61.17		**A**	2.70	2.800	0.038	67.61
	N	0.68	0.702	0.010	67.54		**N**	4.15	4.035	0.055	71.99
	L	0.96	0.846	0.012	79.04		**L**	2.71	2.234	0.031	84.84
	M	0.23	0.226	0.003	70.28		**M**	3.04	2.664	0.037	79.88
A+	**O**	5.83	6.387	0.084	63.90	**A-**	**O**	2.54	2.469	0.034	71.97
	C	3.38	4.127	0.056	57.32		**C**	1.13	1.165	0.016	68.16
	E	12.60	14.474	0.171	60.92		**E**	3.18	3.055	0.042	72.78
	A	3.59	4.291	0.058	58.49		**A**	0.72	0.863	0.012	58.48
	N	4.55	5.286	0.070	60.31		**N**	2.39	2.713	0.037	61.70
	L	5.30	4.657	0.062	79.67		**L**	2.46	2.032	0.028	84.73
	M	3.06	2.957	0.041	72.41		**M**	2.16	1.914	0.027	78.85
N+	**O**	0.95	1.035	0.015	64.41	**N-**	**O**	9.98	10.393	0.129	67.19
	C	3.01	2.935	0.040	71.80		**C**	1.42	1.539	0.022	64.37
	E	2.11	2.317	0.032	63.80		**E**	22.89	27.385	0.281	58.50
	A	0.43	0.584	0.008	51.77		**A**	6.32	7.617	0.098	58.04
	N	1.63	1.542	0.022	73.97		**N**	2.08	2.393	0.033	60.78
	L	4.34	4.077	0.055	74.56		**L**	2.69	2.193	0.030	85.70
	M	2.47	2.983	0.041	58.09		**M**	3.91	4.545	0.061	60.25

Highlighted measurements indicate $p < 0.05$ for that set. We report the sum of squares due to the source (SS), the F-statistic (F), and effect size (η^2). The degree of freedom is 4 for all between groups and 280 for all within-group measurements.

between-group differences are the ones where models act significantly differently. For example, if one model performs very differently than the others, we expect a high value for the sum of **squares due to the source (SS)** between groups. Consequently, SS (between groups) is generally high for the measurements that overlap with the target personality factor of the corresponding set. In general, we observe the highest difference between groups for perceived extraversion.

We generally observe that the models' perceived human likeness and motion appropriateness are more similar than the perceived personality measurements. This behavior suggests that the influence of motion adjustments

Table 2. Pairwise Comparison of the Models for the Positive Personality Sets

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
	S – H	−.373	.000	.199	.089	−.433	.000	−.164	.266	−.164	.311	.145	.605	.125	.666
E+	**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
	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
A+	**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
	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

We report each sample pair's 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 $0.3 > \Delta$. For a pair $A - B$, the mean difference of the corresponding measurement x is calculated as $mean(A_X) - mean(B_X)$. We report pairs with $\rho < 0.05$ and $\Delta > 0.3$ for at least one measure and leave the unabridged results to our supplementary material.

and using different gesture styles do not dramatically alter the agent's realism. On the other hand, perceived extraversion and openness significantly differ between models for most of the sets, followed by agreeableness and neuroticism. For conscientiousness, the models have the slightest difference in general. The models expressing the high traits achieve significant variance overall.

We depict the significant pairwise differences for all models per personality set and measurement type in Table 2 for the sets focusing on expressing high traits and in Table 3 for the sets focusing on expressing low traits; unabridged pairwise differences, as well as an in-depth analysis of the results, are available in Section A of our supplementary material.

We expect the corresponding personality measurements to be high for the models expressing high traits and low for the models expressing low traits. Specific applications may require isolating the expressed personality factor, while others may ignore correlations between the factors. For example, specific models can achieve high performance in expressing the target traits while influencing the perception of other factors. Such models can be preferred if the only aim is to achieve the desired measurement for one factor. On the other hand, if the target personality must be achieved without affecting the perception of others, we also need to observe the changes in other measurements. For this second case, a successful model would achieve a high difference for the target

Table 3. Pairwise Comparison of the Models for the Negative Personality Sets

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

We report each sample pair's 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 $0.3 > \Delta$. For a pair $A - B$, the mean difference of the corresponding measurement x is calculated as $mean(A_X) - mean(B_X)$. We report pairs with $\rho < 0.05$ and $\Delta > 0.3$ for at least one measure and leave the unabridged results to our supplementary material.

personality factor while having a neutral measurement for the remaining factors. For all successful models, high human likeness and motion appropriateness are desired.

In general, models expressing high traits differ in perceived extraversion and openness; we observe a trend for perceiving openness and extraversion together, which we examine later through correlation analysis. Model P depicts high openness and high extraversion the best. Except for Model H, which has energetic movements unsuitable for representing high conscientiousness, models perform similarly for the perceived conscientiousness. Models F and P perform similarly well for expressing high agreeableness. The models expressing high neuroticism have no significant personality difference; however, Model S has significantly better human likeness and motion appropriateness for this set.

Generally, the personality sets expressing the low polarity traits have fewer significant differences between models. We expect the corresponding personality measurement to be low for the successful models of these sets without lowering human likeness and motion appropriateness. The most significant difference is for the negative extraversion set, where Model F achieves the best performance while Model H performs very poorly. The models expressing low agreeableness are measured very similarly. Modifications for negative conscientiousness and neuroticism primarily influence the perceived openness and extraversion. Model F is the most successful in expressing low neuroticism while having high human likeness and motion appropriateness.

5.2 Mean Difference Analysis

An alternative approach for measuring the model performance is to look at the mean difference achieved by the samples expressing opposing traits, which is the measure used in Sonlu et al. [65]. To this end, we compare the mean measurement differences of the high and low personality sets for each model, reporting the mean difference and the significance of the difference based on Welch's t-test in Table 4.

Models that achieve a high difference between their mean personality scores for the opposing target personality samples can depict a broader range of traits and, therefore, are more successful. We expect this difference to be positive for the factor that adjustments focus on, which is valid for all the models. On the other hand, most

Table 4. Mean OCEAN Score, Human Likeness, and Motion Appropriateness Differences for Each Opposing Personality Sample

	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

Δ_x shows the mean score difference between opposing personality samples, and ρ_x is the adjusted p-value for the significance of the difference, where $x \in \{O, C, E, A, N, L, M\}$. Columns show 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 $0.2 > \Delta$.

of the time, participants perceive common traits together [52], which results in a correlation between different personality factors. Such correlations are prominent, especially in zero-acquaintance personality studies utilizing digital characters [18, 63, 65]. For example, the measured extraversion difference between the means of $E+$ and $E-$ samples using Model F is .76; however, this also results in a .77 difference for the measured openness, which may not be ideal for all use cases. Although Model G achieves less extraversion difference between $E+$ and $E-$ samples (.31), its influence on the other factors is minor, meaning it can express extraversion without influencing the other traits.

For human likeness and motion appropriateness measurements, we expect the mean difference for opposing personality expressions to be low. In other words, expressing the high and low traits of the corresponding factor should not impact the human likeness and motion appropriates; in cases where this difference is significant, the modifications due to expressing a certain polarity harm output quality.

Models F and P achieve the highest difference for the perceived openness, followed by Model H. The lack of an expressive style input negatively influences the co-speech generation system's performance in expressing openness. Model P achieves the highest difference for conscientiousness, followed by Model F. The adjustments that focus on expressing conscientiousness alter mainly LMA Time and Flow; using the generated co-speech animations enables less control over the final gesture speed, which can lower the overall performance when using co-speech gesture generation. The phenomena where conscientiousness is perceived inversely proportional to neuroticism [7] is prominent in all the models. Models F and P achieve the best results of distinguishing extraversion at the cost of a high correlation between the measured factors; Model S performs poorly for this personality dimension, likely as it appears low in extraversion for high and low extraversion samples. Gesturing is highly related to perceived extraversion [55]; consequently, Model G achieves notable differences without influencing other factors.

Model P achieves the highest agreeableness difference, followed by Mode F and Model H. Utilizing the happy style helps achieve relatively high performance for the high agreeableness sample in contrast to Model G, which lacks an expressive style input. Personality-based adjustments strengthen the correlation between perceived

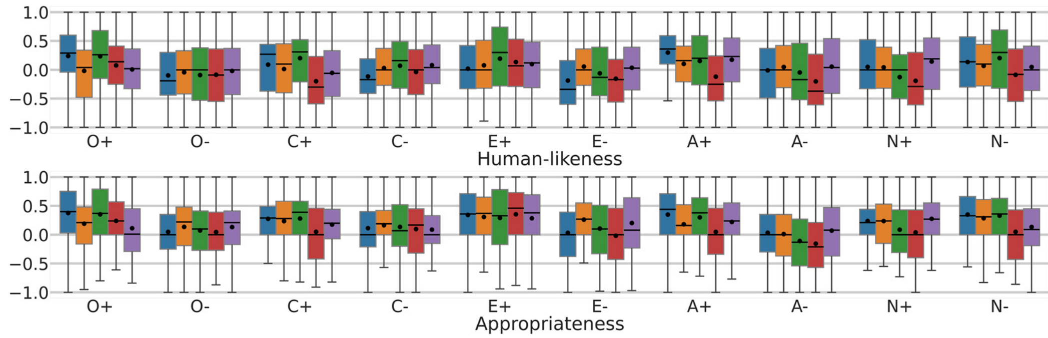


Fig. 6. Box-plot of human likeness and motion appropriateness in the range [−1,1] per personality set and model. Lines depict the median, and the dots show the mean. In each group, the models are ordered from left to right: F, G, P, H, S.

factors. The highest neuroticism difference is achieved by Model F, followed by Models P and G. Correlation between the perceived factors is weaker for Model S, where the lowest neuroticism difference is achieved.

We observe high polarity in human likeness and motion appropriateness of Models F and P, which are generally higher. The most prominent difference in motion appropriateness is for agreeableness modifications done by Model P; conversely, the same changes do not influence the perceived human likeness. The differences in human likeness and motion appropriateness due to extraversion are relatively high for Model H but not prominent for any factor in Models G and S.

Figure 6 shows a box plot of the participants' ratings of each model for human likeness and motion appropriateness. Expressing traits with positive connotations generally improves human likeness and appropriateness. In Section B of our supplementary material, we include a similar visual overview of the personality measurements of the models.

Since the results span many tables, which complicates the comparison of the model performances, we also look at the average distance of the models from an ideal model for each personality set. Suppose an ideal Model I exists that achieves a mean of 1 for the high trait samples and a mean of -1 for the low trait samples when considering the personality factor of interest. This model would also achieve a mean of 0 for the factors we do not intend to express for the corresponding set. For example, Model I would have $O = 1, C = 0, E = 0, A = 0, N = 0$ for $O+$ sample. The Model I would also achieve a mean of 1 for human likeness and motion appropriateness. For the average distance of a Model X from Model I, we calculate their difference per measurement type mean and take its average, which would result in a value in the $[0, 2]$ range. We call this value **"average difference from ideal" (ADFI)**. The smaller the ADFI value is, the more successful the model. **Exclusive-ADFI (eADFI)** considers only the factor of interest for personality measurements, and hence, it gives more importance to human likeness and motion appropriateness. For example, for $O+$ and $O-$ samples, we would only calculate the difference for measurements O, L, M, ignoring any difference in the remaining factors. ADFI and eADFI measurements of each model for each target personality are given in Table 5.

Considering ADFI measurements, Model G performs the best overall. Having appropriate gestures for the speech helps isolate the expression of the desired factor. Model S ranks the second, and Model F ranks the third. Different models take the lead in expressing the opposing personality types; the best-ranking approaches can be utilized for expressing different personalities in applications. For example, considering the perceived openness, an agent that expresses high openness can utilize Model F. In contrast, an agent that expresses low openness can utilize Model H. Model P is the best overall, followed by Model F, based on eADFI measurements. If isolating the personality factor of interest is not desired, personality-based motion adjustments successfully portray the desired traits. Model H ranks last for both overall measurements but can be preferred for low openness. Our

Table 5. ADFI and eADFI Measurements for Each Model Per Personality Sample and the Corresponding Ranks among All Models

Meas.	Model	Personality Sample										Overall
		O+	O-	C+	C-	E+	E-	A+	A-	N+	N-	
ADFI	F	0.481 (1)	0.466 (4)	0.431 (1)	0.505 (5)	0.525 (5)	0.462 (4)	0.466 (1)	0.489 (4)	0.461 (2)	0.404 (2)	0.469 (3)
	G	0.492 (2)	0.459 (2)	0.452 (2)	0.454 (2)	0.474 (2)	0.453 (2)	0.471 (3)	0.450 (2)	0.472 (3)	0.392 (1)	0.457 (1)
	P	0.495 (3)	0.464 (3)	0.506 (4)	0.474 (4)	0.497 (4)	0.425 (1)	0.509 (5)	0.458 (3)	0.475 (4)	0.549 (4)	0.485 (4)
	H	0.496 (4)	0.456 (1)	0.553 (5)	0.451 (1)	0.476 (3)	0.516 (5)	0.507 (4)	0.497 (5)	0.481 (5)	0.552 (5)	0.499 (5)
	S	0.519 (5)	0.475 (5)	0.475 (3)	0.473 (3)	0.464 (1)	0.458 (3)	0.469 (2)	0.448 (1)	0.417 (1)	0.452 (3)	0.465 (2)
eADFI	F	0.693 (2)	0.956 (3)	0.808 (2)	0.975 (5)	0.783 (3)	0.894 (3)	0.697 (1)	0.838 (2)	0.837 (3)	0.697 (1)	0.818 (2)
	G	0.929 (4)	0.911 (2)	0.859 (3)	0.924 (3)	0.841 (4)	0.821 (1)	0.883 (4)	0.867 (3)	0.831 (2)	0.772 (3)	0.864 (3)
	P	0.623 (1)	1.022 (4)	0.751 (1)	0.890 (1)	0.638 (1)	0.927 (4)	0.717 (2)	0.900 (4)	0.919 (4)	0.716 (2)	0.810 (1)
	H	0.771 (3)	1.025 (5)	1.047 (5)	0.923 (2)	0.689 (2)	1.097 (5)	0.972 (5)	0.961 (5)	0.925 (5)	0.957 (5)	0.937 (5)
	S	0.952 (5)	0.903 (1)	0.891 (4)	0.941 (4)	0.888 (5)	0.853 (2)	0.816 (3)	0.825 (1)	0.808 (1)	0.851 (4)	0.873 (4)

For each personality set, the first and second ranking models are highlighted. The overall column shows the average of all personality samples for that model.

results show that each approach has advantages and disadvantages, and a system that aims to portray different personality traits successfully should adopt different strategies based on the requirements.

Section C of our supplementary material includes Principal Component Analysis of the results and Pearson and Kendall correlation between the measurements for each model to observe the related factors in more detail. In general, there is a moderate correlation between human likeness and motion appropriateness for all models. Openness, agreeableness, and extraversion strongly correlate in models that utilize motion adjustments. We observe a strong inverse correlation between conscientiousness and neuroticism, as well as conscientiousness and agreeableness in all models. The correlation between openness and extraversion is prominent in all models. Interestingly, perceived neuroticism has little influence on human likeness in Models G and S; this could mean these models can produce more human-like neurotic animations.

6 Conclusion

With the increasing popularity of data-driven approaches, many applications utilize deep architectures to generate animation for virtual characters. Although existing work focuses on personality expression in heuristics-based systems, automatic animation generation's influence on conversational agents' perceived personality requires further analysis. We can utilize different synthetic animations in agent motion; however, using co-speech gestures is more meaningful for conversational agents to drive behaviors.

In this article, we analyze the influence of co-speech gestures on the apparent personality of virtual characters. We propose a system to utilize data-driven co-speech gestures in an existing multi-modal personality expression framework. We compare five different approaches for expressing the high and low polarities of each personality dimension of the Five-Factor model: using co-speech gestures only with neutral, happy, and sad styles, using personality-specific motion adjustments on the same repeated animation, and using the same motion adjustments on the co-speech gestures generated with the neutral style. We show that different approaches yield the best performance for different target personalities. Specific approaches can disturb the human likeness of the animations while enhancing the personality expression. Combining co-speech gestures with personality adjustments works better for expressing high openness, conscientiousness, and agreeableness. Co-speech gestures with a neutral style are more appropriate for low neuroticism. Motion adjustments on the same repeated animation help better control the overall look, which is more suitable for expressing low extraversion. The happy and sad styles help better isolate the expressed factors but fail to deliver a high difference for the opposing personality types in general. Combining co-speech gestures with motion adjustments yields a solid difference between the high and low trait samples in general, taking advantage of both approaches at the cost of slightly reduced human likeness and motion appropriateness. On the other hand, using only co-speech gestures generally benefits human likeness and motion appropriateness, yet it fails to deliver a solid personality difference for certain factors.

The resulting system can be used in automated conversational agents where an input dialogue is to be performed by characters of different personalities. Intelligent assistants, virtual tutors, and video game characters can benefit from such personality-enhanced co-speech gestures. The results of our experiments inspire further research into the personality implications of co-speech gestures. We analyzed the influence of heuristics-based adjustments on co-speech gestures; however, such adjustments can also be data-driven. Integrating different personalities into an end-to-end co-speech gesture generation system is one of the possibilities for future work. Currently, the system does not recognize the semantics of gestures; the influence of different meaning categories of gestures on apparent personality can be further researched. For example, gesturing increases extroversion, but what if gestures with negative semantics are used only? The timing of different gestures is also essential. Waving while speaking greeting words can have a different impact than waving while a different agent speaks, which requires further analysis. Currently, the gestures are only used while the agent speaks, and the same idle animation is utilized while the agent listens. Using listening animations that follow the speech of the other agent can improve the realism and personality expression. For example, an agent that looks directly at the speaker would appear different than an agent that looks around while the other is speaking. Although in this work, we show that different approaches are suitable for portraying different personality factors, a careful study of our results could enable a more advanced model that is always successful. We hope this work will be a step toward more expressive and human-like virtual characters.

Acknowledgment

We are grateful to Lori Russell Dağ for proofreading the manuscript.

References

[1] Kfir Aberman, Yijia Weng, Dani Lischinski, Daniel Cohen-Or, and Baoquan Chen. 2020. Unpaired motion style transfer from video to animation. *ACM Transactions on Graphics* 39, 4(2020), Article 64, 12 pages.

[2] Chaitanya Ahuja, Dong Won Lee, and Louis-Philippe Morency. 2022. Low-resource adaptation for personalized co-speech gesture generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '22)*. IEEE, 20566–20576.

[3] Chaitanya Ahuja, Dong Won Lee, Yukiko I Nakano, and Louis-Philippe Morency. 2020. Style transfer for co-speech gesture animation: A multi-speaker conditional-mixture approach. In *Proceedings of the16th European Conference onComputer Vision (ECCV '20*). Lecture Notes in Computer Science, Vol. 12363. Springer Nature, 248–265.

[4] Simon Alexanderson, Gustav Eje Henter, Taras Kucherenko, and Jonas Beskow. 2020. Style-controllable speech-driven gesture synthesis using normalising flows. *Computer Graphics Forum* 39, 2 (2020), 487–496.

[5] Tenglong Ao, Qingzhe Gao, Yuke Lou, Baoquan Chen, and Libin Liu. 2022. Rhythmic gesticulator: Rhythm-aware co-speech gesture synthesis with hierarchical neural embeddings. *ACM Transactions on Graphics* 41, 6 (2022), Article 209, 19 pages.

[6] Ali Arya, Lisa N. Jefferies, James T. Enns, and Steve DiPaola. 2006. Facial actions as visual cues for personality. *Computer Animation and Virtual Worlds* 17, 3–4 (2006), 371–382.

[7] Nadin Beckmann, Robert E. Wood, and Amirali Minbashian. 2010. It depends how you look at it: On the relationship between neuroticism and conscientiousness at the within-and the between-person levels of analysis. *Journal of Research in Personality* 44, 5 (2010), 593–601.

[8] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300.

[9] Tamas Bereczkei and Norbert Mesko. 2006. Hair length, facial attractiveness, personality attribution: A multiple fitness model of hairdressing. *Review of Psychology* 13, 1 (2006), 35–42.

[10] Armin Bruderlin, Chor Guan Teo, and Tom Calvert. 1994. Procedural movement for articulated figure animation. *Computers & Graphics* 18, 4 (1994), 453–461.

[11] Justine Cassell, Matthew Stone, Brett Douville, Scott Prevost, Brett Achorn, Mark Steedman, Norm Badler, and Catherine Pelachaud. 2019. Modeling the interaction between speech and gesture. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*. Routledge, 153–158.

[12] Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. 2017. Multimodal human-human-robot interactions (MHHRI) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing* 10, 4 (2017), 484–497.

[13] Yujin Chai, Tianjia Shao, Yanlin Weng, and Kun Zhou. 2024. Personalized audio-driven 3D Facial animation via style-content disentanglement. *IEEE Transactions on Visualization and Computer Graphics* 30, 3 (2024), 1803–1820.

[14] Haoyu Chen, Henglin Shi, Xin Liu, Xiaobai Li, and Guoying Zhao. 2023. SMG: A micro-gesture dataset towards spontaneous body gestures for emotional stress state analysis. *International Journal of Computer Vision* 131, 6 (2023), 1346–1366. DOI: https://doi.org/10.1007/S11263-023-01761-6

[15] Anna Deichler, Siyang Wang, Simon Alexanderson, and Jonas Beskow. 2022. Towards context-aware human-like pointing gestures with RL motion imitation. In *Proceedings of the Context-Awareness in Human-Robot Interaction: Approaches and Challenges, Workshop at 2022 ACM/IEEE International Conference on Human-Robot Interaction*. KTH Royal Institute of Technology Publications, 1–6.

[16] Daniel J. Densley and Philip J. Willis. 1997. Emotional posturing: A method towards achieving emotional figure animation. In *Proceedings of the Computer Animation (CA '97)*. IEEE, 8–14.

[17] Dario Dotti, Mirela Popa, and Stylianos Asteriadis. 2018. Behavior and Personality Analysis in a nonsocial context Dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '18)*. IEEE, 2354–2362.

[18] Funda Durupinar, Mubbasir Kapadia, Susan Deutsch, Michael Neff, and Norman I. Badler. 2016. PERFORM: Perceptual approach for adding OCEAN personality to human motion using laban movement analysis. *ACM Transactions on Graphics* 36 (2016), Article 6, 16 pages.

[19] Paul Ekman. 1957. A methodological discussion of nonverbal behavior. *The Journal of Psychology* 43, 1 (1957), 141–149.

[20] Paul Ekman and Wallace V. Friesen. 1969. Nonverbal leakage and clues to deception. *Psychiatry* 32, 1 (1969), 88–106.

[21] Emiliana. 2019. BVH Tools. Retrieved from https://assetstore.unity.com/packages/tools/animation/bvh-tools-144728

[22] S. Ali Etemad and Ali Arya. 2016. Expert-driven perceptual features for modeling style and affect in human motion. *IEEE Transactions on Human-Machine Systems* 46, 4 (2016), 534–545. DOI: https://doi.org/10.1109/THMS.2016.2537760

[23] Ylva Ferstl, Michael Neff, and Rachel McDonnell. 2021. ExpressGesture: Expressive gesture generation from speech through database matching. *Computer Animation and Virtual Worlds* 32, 3–4 (2021), e2016.

[24] Rong Gao, Xin Liu, Bohao Xing, Zitong Yu, Bjorn W. Schuller, and Heikki Kälviäinen. 2024. Identity-free artificial emotional intelligence via micro-gesture understanding. arXiv:2405.13206 [cs.CV]. Retrieved from https://arxiv.org/abs/2405.13206

[25] Saeed Ghorbani, Ylva Ferstl, Daniel Holden, Nikolaus F. Troje, and Marc-André Carbonneau. 2023. ZeroEGGS: Zero-shot example-based gesture generation from speech. *Computer Graphics Forum* 42, 1 (2023), 206–216. DOI: https://doi.org/10.1111/cgf.14734

[26] Saeed Ghorbani, Kimia Mahdaviani, Anne Thaler, Konrad Kording, Douglas James Cook, Gunnar Blohm, and Nikolaus F. Troje. 2021. MoVi: A large multi-purpose human motion and video dataset. *PLoS One* 16, 6 (2021), e0253157.

[27] Tom Giraud, Florian Focone, Virginie Demulier, Jean Claude Martin, and Brice Isableu. 2015. Perception of emotion and personality through full-body movement qualities: A sport coach case study. *ACM Transactions on Applied Perception* 13, 1 (Oct. 2015), Article 2, 27 pages. DOI: https://doi.org/10.1145/2791294

[28] Ikhsanul Habibie, Mohamed Elgharib, Kripasindhu Sarkar, Ahsan Abdullah, Simbarashe Nyatsanga, Michael Neff, and Christian Theobalt. 2022. A motion matching-based framework for controllable gesture synthesis from speech. In *ACM SIGGRAPH Conference Proceedings*. ACM, New York, NY, Article 46, 9 pages.

[29] Ying Hu, Connor J. Parde, Matthew Q. Hill, Naureen Mahmood, and Alice J. O'Toole. 2018. First impressions of personality traits from body shapes. *Psychological Science* 29, 12 (2018), 1969–1983.

[30] Jennifer Hyde, Elizabeth J. Carter, Sara Kiesler, and Jessica K. Hodgins. 2016. Evaluating Animated Characters: Facial Motion Magnitude Influences Personality Perceptions. *ACM Transactions on Applied Perception* 13, 2 (Feb. 2016), Article 8, 17 pages. DOI: https://doi.org/10.1145/2851499

[31] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. 2013. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 7 (2013), 1325–1339.

[32] David Antonio Gomez Jauregui, Ferran Argelaguet, Anne-Hélène Olivier, Maud Marchal, Franck Multon, and Anatole Lecuyer. 2014. Toward "pseudo-haptic avatars": Modifying the visual animation of self-avatar can simulate the perception of weight lifting. *IEEE Transactions on Visualization and Computer Graphics* 20, 4 (2014), 654–661.

[33] Rune Skovbo Johansen. 2009. *Automated Semi-Procedural Animation For Character Locomotion*. Master's thesis. Department of Information and Media Studies, Aarhus University.

[34] Kacper Kania, Marek Kowalski, and Tomasz Trzciński. 2021. TrajeVAE–controllable human motion generation from trajectories. arXiv:2104.00351. Retrieved from https://arxiv.org/abs/2104.00351.

[35] Dacher Keltner. 1996. Facial expressions of emotion and personality. In *Handbook of Emotion, Adult Development, and Aging*. Carol Magai and Susan H. McFadden (Eds.), Elsevier, 385–401.

[36] Sophie Kenny, Naureen Mahmood, Claire Honda, Michael J. Black, and Nikolaus F. Troje. 2019. Perceptual Effects of Inconsistency in Human Animations. *ACM Transactions on Applied Perception* 16, 1 (Feb. 2019), Article 2, 18 pages. DOI: https://doi.org/10.1145/3301411

[37] Hye Ji Kim, Michael Neff, and Sung-Hee Lee. 2022. The perceptual consistency and association of the LMA effort elements. *ACM Transactions on Applied Perception* 19, 1 (Jan. 2022), Article 1, 17 pages. DOI: https://doi.org/10.1145/3473041

[38] Robert M. Krauss. 1998. Why do we gesture when we speak? *Current Directions in Psychological Science* 7, 2 (1998), 54–54.

[39] Robert M. Krauss, Yihsiu Chen, and Purnima Chawla. 1996. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In *Proceedings of the Advances in Experimental Social Psychology*. Vol. 28. Elsevier, 389–450.

[40] Taras Kucherenko, Patrik Jonell, Youngwoo Yoon, Pieter Wolfert, and Gustav Eje Henter. 2021. A large, crowdsourced evaluation of gesture generation systems on common data: The GENEA Challenge 2020. In *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)*. ACM, New York, NY, 11–21.

[41] Deng Li, Xin Liu, Bohao Xing, Baiqiang Xia, Yuan Zong, Bihan Wen, and Heikki Kälviäinen. 2024. EALD-MLLM: Emotion analysis in long-sequential and de-identity videos with multi-modal large language model. arXiv:2405.00574 [cs.CV]. Retrieved from https://arxiv.org/abs/2405.00574

[42] Deng Li, Bohao Xing, and Xin Liu. 2024. Enhancing micro gesture recognition for emotion understanding via context-aware visual-text contrastive learning. *IEEE Signal Processing Letters* 31 (2024), 1309–1313. DOI: https://doi.org/10.1109/LSP.2024.3396656

[43] Jingyi Li, Michelle X Zhou, Huahai Yang, and Gloria Mark. 2017. Confiding in and listening to virtual agents: The effect of personality. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI '17)*. ACM, New York, NY, 275–286.

[44] Yuanzhi Liang, Qianyu Feng, Linchao Zhu, Li Hu, Pan Pan, and Yi Yang. 2022. SEEG: Semantic energized co-speech gesture generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '22)*. IEEE, 10473–10482.

[45] Xin Liu, Henglin Shi, Haoyu Chen, Zitong Yu, Xiaobai Li, and Guoying Zhao. 2021. iMiGUE: An identity-free video dataset for micro-gesture understanding and emotion analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '21)*. IEEE, 10626–10637. DOI: https://doi.org/10.1109/CVPR46437.2021.01049

[46] IBM. 2015. IBM Watson API. Retrieved May 01, 2023 from https://www.ibm.com/watson

[47] Oculus. 2021. Oculus Lipsync. Retrieved August 14, 2022 from https://developer.oculus.com/downloads/package/oculus-lipsync-unity/

[48] Rachel McDonnell, Sophie Jörg, Jessica K. Hodgins, Fiona Newell, and Carol O'sullivan. 2009. Evaluating the Effect of Motion and Body Shape on the Perceived Sex of Virtual Characters. *ACM Transactions on Applied Perception* 5, 4 (Feb. 2009), Article 20, 14 pages. DOI: https://doi.org/10.1145/1462048.1462051

[49] Albert Mehrabian. 1968. Some referents and measures of nonverbal behavior. *Behavior Research Methods & Instrumentation* 1, 6 (1968), 203–207.

[50] Jianyuan Min, Yen-Lin Chen, and Jinxiang Chai. 2009. Interactive generation of human animation with deformable motion models. *ACM Transactions on Graphics* 29, 1 (2009), Article 9, 12 pages.

[51] Franck Multon, Laure France, Marie-Paule Cani-Gascuel, and Giles Debunne. 1999. Computer animation of human walking: A survey. *The Journal of Visualization and Computer Animation* 10, 1 (1999), 39–54.

[52] Janek Musek. 2007. A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality* 41, 6 (2007), 1213–1233.

[53] Laura P. Naumann, Simine Vazire, Peter J. Rentfrow, and Samuel D. Gosling. 2009. Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin* 35, 12 (2009), 1661–1671.

[54] Simbarashe Nyatsanga, Taras Kucherenko, Chaitanya Ahuja, Gustav Eje Henter, and Michael Neff. 2023. A comprehensive review of data-driven co-speech gesture generation. *Computer Graphics Forum* 42, 2 (2023), 569–596. DOI: https://doi.org/10.1111/CGF.14776

[55] Sinead O'Carroll, Elena Nicoladis, and Lisa Smithson. 2015. The effect of extroversion on communication: Evidence from an interlocutor visibility manipulation. *Speech Communication* 69 (2015), 1–8.

[56] S Mark Pancer and James R Meindl. 1978. Length of hair and beardedness as determinants of personality impressions. *Perceptual and Motor Skills* 46, 3_suppl (1978), 1328–1330.

[57] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. 2018. DeepMimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics* 37, 4 (2018), Article 143, 14 pages.

[58] Tim Polzehl, Sebastian Möller, and Florian Metze. 2010. Automatically assessing personality from speech. In *Proceedings of the IEEE Fourth International Conference on Semantic Computing*. IEEE, 134–140.

[59] Tanmay Randhavane, Aniket Bera, Kyra Kapsaskis, Rahul Sheth, Kurt Gray, and Dinesh Manocha. 2019. EVA: Generating emotional behavior of virtual agents using expressive features of gait and gaze. In *Proceedings of the ACM Symposium on Applied Perception (SAP '19)*. ACM, New York, NY, 1–10.

[60] Klaus R Scherer. 1978. Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology* 8, 4 (1978), 467–487.

[61] Henglin Shi, Wei Peng, Haoyu Chen, Xin Liu, and Guoying Zhao. 2022. Multiscale 3D-shift graph convolution network for emotion recognition from human actions. *IEEE Intelligent Systems* 37, 4 (2022), 103–110. DOI: https://doi.org/10.1109/MIS.2022.3147585

[62] Harrison Jesse Smith, Chen Cao, Michael Neff, and Yingying Wang. 2019. Efficient neural networks for real-time motion style transfer. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 2, 2 (2019), 1–17.

[63] Harrison Jesse Smith and Michael Neff. 2017. Understanding the Impact of Animated Gesture Performance on Personality Perceptions. *ACM Transactions on Graphics* 36, 4 (2017), Article 59, 12 pages.

[64] Sinan Sonlu, Yalim Doğan, Arçin Ülkü Ergüzen, Musa Ege Ünalan, Serkan Demirci, Funda Durupinar, and Uğur Güdükbay. 2024. Towards understanding personality expression via body motion. In *Proceedings of the IEEE VR 2024 2nd Annual Workshop on Multi-modal Affective and Social Behavior Analysis and Synthesis in Extended Reality (MASSXR '24)*. IEEE, 628–631.

[65] Sinan Sonlu, Uğur Güdükbay, and Funda Durupinar. 2021. A conversational agent framework with multi-modal personality expression. *ACM Transactions on Graphics* 40, 1 (2021), Article 7, 16 pages.

[66] Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H Bermano. 2022. Human motion diffusion model. arXiv:2209.14916. Retrieved from https://arxiv.org/abs/2209.14916.

[67] Nguyen Tan Viet Tuyen and Oya Celiktutan. 2022. Context-aware human behaviour forecasting in dyadic interactions. In *Understanding Social Behavior in Dyadic and Small Group Interactions (Proceedings of Machine Learning Research),* Vol. 173. Cristina Palmero, Julio C. S. Jacques Junior, Albert Clapés, Isabelle Guyon, Wei-Wei Tu, Thomas B. Moeslund, and Sergio Escalera (Eds.), ML Research Press, Online, 88–106. Retrieved from https://proceedings.mlr.press/v173/tuyen22a.html

[68] Munetoshi Unuma, Ken Anjyo, and Ryozo Takeuchi. 1995. Fourier principles for emotion-based human figure animation. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*. ACM, New York, NY, 91–96.

[69] Yingying Wang, Jean E. Fox Tree, Marilyn Walker, and Michael Neff. 2016. Assessing the impact of hand motion on virtual character personality. *ACM Transactions on Applied Perception* 13, 2 (Mar. 2016), Article 9, 23 pages. DOI: https://doi.org/10.1145/2874357

[70] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. 2022. MotionDiffuse: Text-driven human motion generation with diffusion model. arXiv:2208.15001. Retrieved from https://arxiv.org/abs/2208.15001.

[71] Katja Zibrek, Sean Martin, and Rachel McDonnell. 2019. Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Transactions on Applied Perception* 16, 3 (2019), 1–19.

Received 17 June 2023; revised 13 August 2024; accepted 3 September 2024