

Technical Section

Personality perception in human videos altered by motion transfer networks

Ayda Yurtoğlu¹, Sinan Sonlu¹, Yalım Doğan¹, Uğur Güdükbay^{*}

Department of Computer Engineering, Bilkent University, 06800, Ankara, Turkey

ARTICLE INFO

Dataset link: <https://paperswithcode.com/dataset/ted-talks>

Keywords:

Personality perception
 Five-Factor model
 Motion transfer networks
 Optical flow
 Motion cues
 Physical appearance

ABSTRACT

The successful portrayal of personality in digital characters improves communication and immersion. Current research focuses on expressing personality through modifying animations using heuristic rules or data-driven models. While studies suggest motion style highly influences the apparent personality, the role of appearance can be similarly essential. This work analyzes the influence of movement and appearance on the perceived personality of short videos altered by motion transfer networks. We label the personalities in conference video clips with a user study to determine the samples that best represent the Five-Factor model's high, neutral, and low traits. We alter these videos using the Thin-Plate Spline Motion Model, utilizing the selected samples as the source and driving inputs. We follow five different cases to study the influence of motion and appearance on personality perception. Our comparative study reveals that motion and appearance influence different factors: motion strongly affects perceived extraversion, and appearance helps convey agreeableness and neuroticism.

1. Introduction

Personality investigates the qualities that form the individual. The Five-Factor model [1] examines personality under five orthogonal dimensions: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. An accurate portrayal of these factors helps improve immersion and communication [2–4]; applications utilize motion [5] and physical appearance [6] features to convey different personalities. While both approaches are well studied in virtual characters separately, the interaction between appearance and movement remains to be explored, especially in videos altered by motion transfer networks that enable a practical way for character animation. This work analyzes the impact of appearance and motion in conveying different personalities of the Five-Factor model in videos altered by motion transfer networks.

Expressive systems utilize a high-level understanding of human motion to define heuristics-based rules that control the virtual human's animation and personality [2]. Such high-level features also apply to motion [7] and personality [8] analysis. Appearance-based features, on the other hand, vary dramatically: Hair color [9], hair length [10], facial shape, skin texture, and viewing angle [11], body shape [12,13], clothing [14], and eye size [15] can influence apparent personality. Applications generally focus on expressing personality in predetermined characters and thus do not aim to alter the appearance. In an animated film, different scenes may require the same character to express different personality traits without a change in physical

appearance; however, the general appearance is highly influential on perceived personality [6].

Suppose we have two characters of different appearances. Would adopting the same motion expressing extraversion similarly influence their apparent personality? Which would be more influential if the appearance expresses extraversion and the movements express the opposite? These questions inspired us to investigate the interaction of appearance- and motion-based cues in videos conveying personality.

With the rise of deep generative models [16–18], the future of CGI can be data-driven. While the ownership of artificial creativity is still in debate [19], studies utilize data-driven methods to synthesize the motion [20,21] and the looks of the virtual characters [22]. Most of our digital characters could adopt such generative methods in the future, and well representing different personalities in virtual human characters would be essential for a more diverse digital world; we analyze the capabilities of a motion transfer network to modify human videos to express different personalities as a step towards this goal.

We label the apparent personalities of the individuals in the TED-talks dataset [23] using Ten-Item Personality Inventory [24] in a user study. We pick the samples representing high, neutral, and low traits for each of the five personality dimensions. We use the Thin-Plate Spline Motion Model (TPS) [25] to generate new sequences with the selected samples, using them as the source and driving inputs. The output videos have the source's appearance and the driving sample's motion. We examine five different cases to examine the influence of different input combinations.

* Corresponding author.

E-mail addresses: ayda.yurtoglu@ug.bilkent.edu.tr (A. Yurtoğlu), sinan.sonlu@bilkent.edu.tr (S. Sonlu), yalim.dogan@bilkent.edu.tr (Y. Doğan), gudukbay@cs.bilkent.edu.tr (U. Güdükbay).¹ Joint first authors.

We list our research questions as follows:

- Can motion transfer networks alter personality perception in human videos?
- Do appearance and motion-based cues similarly impact personality expression?
- Do adjustments made by motion transfer networks influence each personality factor similarly?

This work's contribution is an extensive analysis of TPS's performance in portraying personality through utilizing different samples as the source and driving inputs. Our framework, data, and user study results are available on our GitHub repository² for further studies. The results suggest that using the corresponding high-trait samples as driving input helps portray extraversion and openness; using them as the source input helps express agreeableness and neuroticism. The influence of motion transfer on perceived conscientiousness is limited.

2. Related work

Theories emphasize different elements as the source of personality [26]; biological approaches focus on the link between genetics and apparent traits [27], behavioral approaches explain the causes through interaction with the environment [28], and psychodynamic approaches highlight the unconscious mind [29]. Although personality models investigate similar human traits, their categorizations vary [30]. Among the many models used in psychology, the Five-Factor model [1] is popular in computer science due to its orthogonal dimensions.

Both static cues, such as hair and clothing style, and dynamic cues, like facial expression and posture, signal personality-relevant information [6,31]. People's judgment of others' personalities depends on such cues in first encounters. Research indicates peer and self-reported personalities mostly agree for extraversion [32] and vary in accuracy for other factors [33–35]. Accurate judgment depends on both parties; the peer should observe the cues correctly, and the self should express them clearly [36]. In computer-generated imagery, faithful representation of the cues is essential for expressing the desired personality. The successful portrayal of different personality traits in virtual humans can build trust [37], improve social presence [38] and evoke empathy [39]. Consequently, many applications that involve human characters desire an appropriate portrayal of the personality.

2.1. Expressing personality

Studies utilize expressive motion to create more human-like virtual characters. Appropriate gesturing improves realism and engagement [40]. Data-driven models can generate hand gestures accompanying input speech [41]. Appropriate use of facial expressions [42] and nonverbal communication cues [43] induce empathy in users. Facial expressions reveal the individual's feelings with high accuracy [44], and the magnitude of facial motion in animated characters influences personality perception [45]; consequently, personality recognition systems often input facial [46] and body related [47] features together. Movement qualities such as smoothness, energy, and impulsivity correlate with apparent personality [48]. Body shape and motion style influence the perceived sex of virtual characters [49], causing users to behave differently to them [50].

Even with minimal information, motion indicators accurately predict personality [51], inspiring studies to utilize gesturing [3,52] and hand motion [53] in expressive animation. High-level motion features control the apparent personality [2] and are useful in motion classification and emotion analysis [54–57]. Utilizing personality-specific dialogues, vocal features, and facial expressions improves the distinction of opposite personalities in conversational agents [4]; speech and animation realism influence the perceived personality [58].

Studies alter 3D face models [59], use automatic gaze animation [60] and synthetic body poses [61] to express different personalities. Similar techniques apply to interactive humanoid robots [62]. Imperfections in virtual characters can cause an uncanny effect [63] when they are very close to real but have noticeable divergence [64]; this results in decreased trustworthiness [65] and empathy [66] towards expressive virtual characters. Positive traits like high agreeableness and emotional stability can weaken this negative effect [67]. A recent study reveals that virtual characters' appearance and animation realism influence the perceived social presence and affinity while portraying different emotional facial expressions [68]. Unlike existing research, this work examines the personality expression in human videos altered by motion transfer networks.

2.2. Generating and altering human videos

There is an increasing interest in generating and altering human videos using data-driven models. Supervised methods in this field utilize 3D models [69–73], domain labels [74], semantic segmentations [75], and landmarks [76–80]. Generalizing these methods to different object categories is challenging and often requires fine-tuning [77] or retraining [78]. Unsupervised image animation techniques address the limitations of the supervised methods and allow arbitrary subjects to be animated without prior knowledge, using a source and a driving input. For instance, Monkey-Net [81] acquires motion information via self-learned key points and applies the corresponding trajectories to animate the source image. First Order Motion Model (FOMM) [82] transfers complex motions to arbitrary objects by combining self-learned key points with local affine transformations, applying first-order Taylor expansions to approximate the motion near each key point; however, as local affine transformations are linear, they create limitations for objects that do not move linearly.

Motion Representations for Articulated Animation (MRAA) provides a warp-based method to identify object parts and extract their motion using PCA-based estimation [23]. MRAA predicts the background motion separately to overcome the limitations of FOMM. These warp-based methods use optical flow estimators that warp the source image features to the driving video. Recycle-GAN [83] utilizes a conditional GAN focusing on the video-to-video translation of arbitrary objects; unlike FOMM and MRAA, Recycle-GAN uses warp-free synthesis. MoCoGAN [84] is an unconditional GAN that controls the content and movement of the input videos. To address non-linear movements of the human body, TransMoMo [85] is designed explicitly for representing articulated motion and disentanglements. Unlike the other data-driven models, Transformation-Synthesis Network (TS-Net) [86] has a dual branch structure: warp-based transformation branch and warp-free synthesis branch, which are used together to generate video frames.

We use TPS [25], a warp-based motion transfer framework that addresses the shortcomings of previous unsupervised methods. It can generate decent-quality videos even when a large pose gap exists between the source image and the driving video. It also overcomes the limitations of local affine transformations using TPS motion estimation. It demonstrates a better optical flow performance with a substantial inpainting capability than its ancestors.

3. The method and experiments

We first choose short video clips varying in motion and appearance from a dataset of conference talks. We label the personalities they represent in our first user study. Then, we identify the samples that best represent high, neutral, and low traits for each personality factor; we use them as TPS's source and driving inputs to generate new videos. Our second user study compares TPS's outputs following five cases to examine the influence of motion and appearance on personality perception.

² <https://github.com/sinansonlu/motion-transfer-personality>.

Table 1

Summary of the Five-Factor personality dimensions and the corresponding high and low polarity keywords used in the first experiment.

| Factor | Related traits | High polarity | Low polarity |
|-------------------|---|------------------------------------|----------------------------|
| Openness | The intellectual aspect of the personality, includes traits such as insightfulness, curiosity and self-examination. | Open to new experiences Complex | Conventional Uncreative |
| Conscientiousness | Describes thoughtfulness and goal-directed behavior, conscientious people are organized and mindful of details. | Dependable Self-disciplined | Disorganized Careless |
| Extraversion | Reflects seeking social connections and interaction, extraverts are outgoing and comfortable in social situations. | Extraverted Enthusiastic | Reserved Quiet |
| Agreeableness | Measures the degree of compassion and cooperation, agreeable people feel more empathy and concern for others. | Sympathetic Warm | Critical Quarrelsome |
| Neuroticism | Explains the tendency towards moodiness and emotional instability, neurotic people feel more stress and anxiety. | Anxious Easily upset | Calm Emotionally stable |

3.1. Choosing the samples for personality questionnaire

We manually label the available samples from the TED-Talks dataset [23] based on the presenter's age, gender, race, clothing style (plain or complex), and motion (active or passive). We randomly picked 50 samples of unique label combinations to include in our first user study to better represent different groups and personalities. Among the 50 samples, 29 included a female speaker, 9 represented old age, 13 wore glasses, 17 had long hair, 2 had ties, 10 had facial hair, and 17 wore bright colors. Complete labels of the selected samples are available in our repository. Each sample is a muted sequence of 5–10 s from TED Talks. The videos focus on the presenter and only show the upper part of the body.

3.2. User studies

We run two online user studies³ using our website implemented for this project. Participants could use any device supporting HTML and JavaScript to contribute to the user study. We acquire the participants among university students and their family members through email invitations; participation is entirely voluntary. Participants log in to the website using their email addresses and can resume the experiment anytime.

We did not prevent the users from participating in the second study if they participated in the first one, which enabled us to study both groups individually to analyze the influence of early exposure to the appearance and motion of the samples. Of all the 47 participants of the first study, 40 provided optional demographics (23 male and 17 female) with an average age of 34.45 ± 12.93 . 11 of these users also participated in the second study. The total participant count for the second study is 30; 28 participants provided demographics (21 male and seven female) with an average age of 28.42 ± 11.39 .

3.3. First user study

In the first study, we show the selected 50 videos in random order one by one. Participants rate the individual's personality using sliders on a 7-point Likert scale. Each slider contains descriptive traits for opposing polarities of the personality factors from the Ten-Item Personality Inventory (TIPI) [24], shown in Table 1. We adopt this brief usage of TIPI from PERFORM's comparison study [2]. We map the slider values to integers in the $[-3, 3]$ range, where negative and positive values correspond to low and high traits of the corresponding personality factors, respectively. A value close to zero represents neutral traits.

We use one-way Analysis of Variance (ANOVA) to investigate if the personality means of the samples differ significantly. Each personality factor is the dependent variable, and samples are the independent variables. Table 2 shows samples were rated significantly different

Table 2

ANOVA statistics per personality factor where samples are the independent variable and the participants' ratings are the dependent variable.

| | Source | SS | df | F | η^2 |
|---|----------------|---------|------|----------|----------|
| O | Between groups | 486.34 | 49 | 5.529** | 0.172 |
| | Within groups | 2336.89 | 1302 | – | – |
| C | Between groups | 263.93 | 49 | 3.517** | 0.116 |
| | Within groups | 1994.02 | 1302 | – | – |
| E | Between groups | 1222.64 | 49 | 12.753** | 0.324 |
| | Within groups | 2547.32 | 1302 | – | – |
| A | Between groups | 866.91 | 49 | 8.943** | 0.251 |
| | Within groups | 2575.54 | 1302 | – | – |
| N | Between groups | 609.13 | 49 | 5.589** | 0.173 |
| | Within groups | 2895.49 | 1302 | – | – |

** Statistics marked indicate $p < 0.001$.

for all personality factors. The extraversion variance is the highest, which can be due to this factor being the easiest to perceive. On the other hand, conscientiousness variance is the lowest; this is likely as all the samples are presenters, and thus they score relatively high in conscientiousness.

Among our 50 samples, we choose X_H and X_L as the factors representing each personality factor's high and low polarity, respectively, where $X \in O, C, E, A, N$. We compare each sample pair based on their mean difference in each factor's score using Tukey Honestly Significant Differences (HSD) [87]; for each factor, the pair with the highest mean difference is selected as X_H and X_L . For example, sample 45 and 20 have the highest mean extraversion difference. The mean extraversion of sample 45 is 2.28, so it is selected as E_H , and sample 20 has a mean extraversion of -1.75 , and consequently, it is selected as E_L . For each factor, we also choose a sample to represent the neutral traits X_N , where $X \in O, C, E, A, N$ such that $|Mean(X_H) - Mean(X_N)| + |Mean(X_N) - Mean(X_L)|$ is maximized, where the mean function calculates the mean score for the corresponding factor. We allow a sample to represent multiple factors; for example, sample 9 represents O_L (low openness) and N_N (neutral neuroticism). The index numbers of the samples correspond to their order in our dataset, so we replace them with X_H , X_N , and X_L for the selected samples.

Table 3 shows the pairwise comparisons between the selected samples per factor. Although we are only interested in the matching personality measurement per sample pair, we report all factors' ratings. For example, E_H , E_N , and E_L are selected to represent extraversion, but we also report their differences for remaining factors. The results show that the selected X_H – X_L pairs are most suitable to represent their matching factors but can also have significant differences for other factors. For instance, E_H – E_L has a mean difference of 4.03 in extraversion and a mean difference of 1.54 in conscientiousness. Such correlation between different factors can be observed in similar work [3,4]. We ran the Kaiser–Meyer–Olkin test on our data to see if this correlation was too high; however, the resulting value of .499 indicates that it is unsuitable

³ Bilkent University Ethical Committee for Human Research approves the studies with the decision number 2023_06_09_01.

Table 3

Pairwise comparison of the selected high (X_H), neutral (X_N), and low (X_L) samples, where $X \in \{O, C, E, A, N\}$. The rows show the personality factor that is compared; we report the mean difference and Tukey HSD adjusted p values for each sample pair. Bold values show $p < 0.05$.

| Factor | Sample pair | Mean Diff. | Adj. P | Sample pair | Mean Diff. | Adj. P | Sample pair | Mean Diff. | Adj. P | Sample pair | Mean Diff. | Adj. P | Sample pair | Mean Diff. | Adj. P |
|--------|-------------|--------------|-----------------|-------------|--------------|-----------------|-------------|---------------|-----------------|-------------|---------------|-----------------|-------------|---------------|-----------------|
| O | O_H-O_L | 2.412 | <.001 | C_H-C_L | -0.494 | .999 | E_H-E_L | 1.130 | .573 | A_H-A_L | -0.165 | .999 | N_H-N_L | -0.179 | .999 |
| | O_H-O_N | 1.107 | .549 | C_H-C_N | 0.149 | .999 | E_H-E_N | 0.639 | .999 | A_H-A_N | 0.308 | .999 | N_H-N_N | 1.019 | .795 |
| | O_N-O_L | 1.305 | .191 | C_N-C_L | -0.643 | .999 | E_N-E_L | 0.491 | .999 | A_N-A_L | -0.472 | .999 | N_N-N_L | -1.198 | .387 |
| C | O_H-O_L | -0.223 | .999 | C_H-C_L | 2.077 | <.001 | E_H-E_L | 1.540 | .006 | A_H-A_L | 1.030 | .557 | N_H-N_L | -1.286 | .074 |
| | O_H-O_N | -0.107 | .999 | C_H-C_N | 0.792 | .985 | E_H-E_N | 0.385 | .999 | A_H-A_N | 0.001 | .999 | N_H-N_N | -1.615 | .002 |
| | O_N-O_L | -0.115 | .999 | C_N-C_L | 1.286 | .074 | E_N-E_L | 1.155 | .220 | A_N-A_L | 1.030 | .557 | N_N-N_L | 0.330 | .999 |
| E | O_H-O_L | 1.901 | .001 | C_H-C_L | -0.167 | .999 | E_H-E_L | 4.030 | <.001 | A_H-A_L | 0.134 | .999 | N_H-N_L | -0.786 | .997 |
| | O_H-O_N | 1.893 | .001 | C_H-C_N | -0.381 | .999 | E_H-E_N | 1.694 | .009 | A_H-A_N | 1.577 | .040 | N_H-N_N | -1.635 | .017 |
| | O_N-O_L | 0.008 | .999 | C_N-C_L | 0.214 | .999 | E_N-E_L | 2.336 | <.001 | A_N-A_L | -1.443 | .093 | N_N-N_L | 0.849 | .991 |
| A | O_H-O_L | 1.459 | .096 | C_H-C_L | 1.708 | .013 | E_H-E_L | -0.367 | .999 | A_H-A_L | 3.546 | <.001 | N_H-N_L | 0.321 | .999 |
| | O_H-O_N | -0.429 | .999 | C_H-C_N | 1.530 | .069 | E_H-E_N | -0.726 | .999 | A_H-A_N | 1.615 | .030 | N_H-N_N | 1.530 | .051 |
| | O_N-O_L | 1.887 | .001 | C_N-C_L | 0.179 | .999 | E_N-E_L | 0.358 | .999 | A_N-A_L | 1.931 | .001 | N_N-N_L | -1.209 | .493 |
| N | O_H-O_L | 0.706 | .999 | C_H-C_L | -0.827 | .999 | E_H-E_L | -1.357 | .367 | A_H-A_L | -1.464 | .162 | N_H-N_L | 3.000 | <.001 |
| | O_H-O_N | 0.786 | .999 | C_H-C_N | -0.970 | .980 | E_H-E_N | 0.828 | .999 | A_H-A_N | 0.885 | .996 | N_H-N_N | 1.742 | .017 |
| | O_N-O_L | -0.080 | .999 | C_N-C_L | 0.143 | .999 | E_N-E_L | -2.185 | <.001 | A_N-A_L | -2.349 | <.001 | N_N-N_L | 1.258 | .544 |

Table 4

Motion and appearance combinations used in the second study.

| Case | Output A | | Output B | |
|------|----------|---------|----------|---------|
| | Source | Driving | Source | Driving |
| 1 | Neutral | High | Neutral | Low |
| 2 | High | Neutral | Low | Neutral |
| 3 | Low | High | High | Low |
| 4 | High | High | High | Low |
| 5 | Low | High | Low | Low |

for factor analysis, and hence, the orthogonality of the samples is well captured.

We illustrate the selected high, neutral, and low samples per factor in Fig. 1 with their personality block diagrams. Note that the samples of low extraversion – high neuroticism and low openness – neutral neuroticism are the same; this does not cause a problem for the second study as we focus on each factor individually, and there is no case where samples representing different factors are combined. Although we only report the personality ratings and pairwise comparisons for the selected samples in the article, the complete data and comparisons are available on our public repository for further analysis.

3.4. Altering samples for the second study

We use TPS to alter the selected samples representing high, neutral, and low personality traits. For this task, we utilize different videos as TPS’s source and driving inputs to output the video pairs introduced in Table 4. We form these output pairs for each personality factor. We aim to cover as many of the possible combinations while keeping the sample size manageable. We compare these altered samples to determine which combination represents the high traits better in our second study.

Three cases measure the impact of opposing trait motions on the same source image. We use the samples selected for neutral, high, and low traits as the source input in Cases 1, 4, and 5, respectively; we determine if the source image affects perception when appearance is the same and motions are opposing. Case 2 investigates the effect of appearance where we use the neutral sample’s motion as the driving input, and opposing traits are used as source inputs; we only use the neutral motion and omit the combinations that use high and low samples as the driving input on the opposing sources to keep the focus on the appearance. As the motions are the same for these combinations, we expect a similar motion transfer quality; in contrast, as Cases 1, 4, and 5 cover different source images, the performance of TPS can vary, which could influence perception. We discuss these possible artifacts due to

the mismatch between source and driving inputs in the Limitations section.

In each comparison, we use high and low samples as the source or driving input to keep the stimuli more practical; we do not compare high and neutral driving inputs on the same source or any such combination. Case 3 analyzes whether driving or source input is more dominant when they utilize opposing samples. We find out which combination expresses the high trait better: Low trait appearance with high trait motion or high trait appearance with low trait motion. For each personality factor, we have five cases resulting in 25 output pairs.

3.5. Motion transfer

We utilize the pre-trained version of TPS, which takes an RGB source image and a driving video of 384×384 resolution for motion transfer; the motion of the driving video is transferred to the actor in the source image. The synthesized video has the exact frame count as the driving video. As our samples are videos, we need to choose a video frame to be used as the source image, and the pose in the selected frame determines the quality of the output. In cases where the actor in the source image is blurry or has an abnormal pose, the output suffers.

We follow the recommended alignment method in the original TPS implementation to prevent low-quality results. The authors suggest finding an aligning frame in the driving video with a pose similar to the source image. The alignment is calculated by determining the face landmarks and calculating their 2D convex hull. The landmarks are centered and normalized according to the square root of the area of the convex hull. We take the frame that minimizes the squared error between source image landmarks as the representative frame. The motion transfer is initialized from this frame in the driving video and processed in both directions: towards the beginning and the end. The final videos are stitched accordingly, and the final output is obtained.

Since calculating the representative frame for each pair of videos would be time-consuming, we pre-process all videos using a standard idle pose. We first calculate 3D landmarks of the human body using MediaPipe Pose Landmarker [88] and calculate the 3D convex hull. The frame in each video closest to the idle pose is taken as the representative frame of that video. An example representative frame and its extracted pose can be seen in Fig. 2. During motion transfer, the representative frame of the source video is taken as the source image. The stitching logic in the original work mentioned above is applied as is. An example motion transfer between a source image and a driving video can be seen in Fig. 3. Notice the slight deformation of the face in the source image as they face different directions. Such deformations can reduce the quality of the outputs, influencing the perception of desired traits.

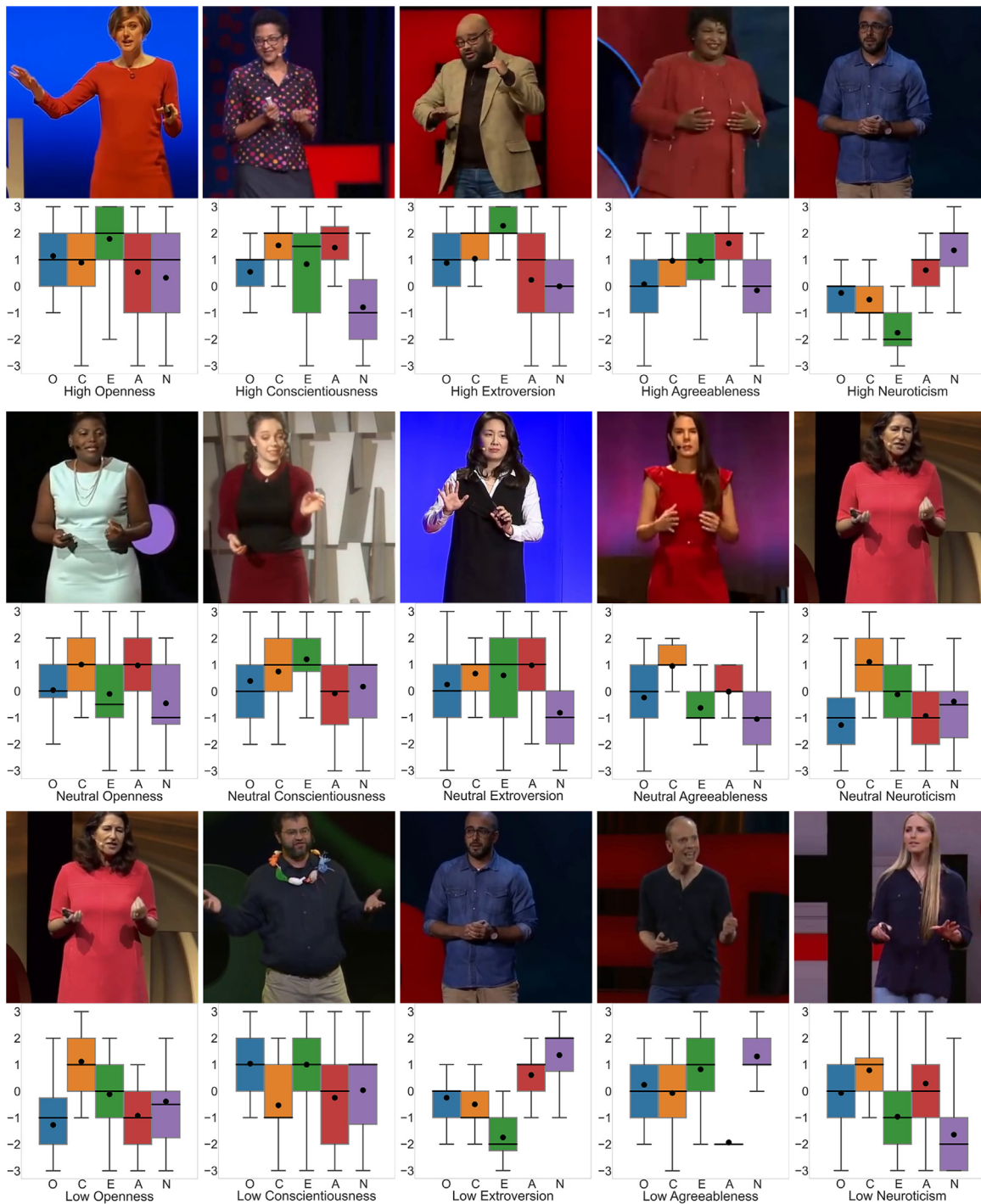


Fig. 1. Selected samples representing high, low, and neutral traits for the corresponding personality factors. Each box plot summarizes the personality scores of the sample below; black lines depict the median, and the black dots show the mean.

3.6. Second user study

The second study compares the altered output pairs introduced in Table 4. In each task, we play the matched output pairs simultaneously. Participants are free to repeat, pause, and rewind the videos. The order and left–right placement of the outputs are randomized. Our first question asks, “Which video contains a more **extraverted**, **enthusiastic** and less **reserved**, **quiet** looking person?”. We replace the keywords written in bold using the traits from Table 1 based on the personality factor we measure. The second question asks, “Which video looks more realistic?”. For each question, the participant selects one of

the buttons indicating *Left*, *Equal*, and *Right*. We require watching the videos at least once before submitting. At the beginning of the study, we show a text briefly explaining the Five-Factor personality traits to the participants. We inform the participants that the samples are modified using software and define a realistic-looking sample as one with no noticeable marks of the modification. In this case, a less realistic sample has artifacts that reveal the modifications on the video.

We map the users’ answers to the integers 0, 1, and 2 for choosing Output A, Equal, or Output B, respectively, to be used as the dependent variable. We compare output pairs of TPS following the combinations of five cases. We only focus on the personality factor that input samples

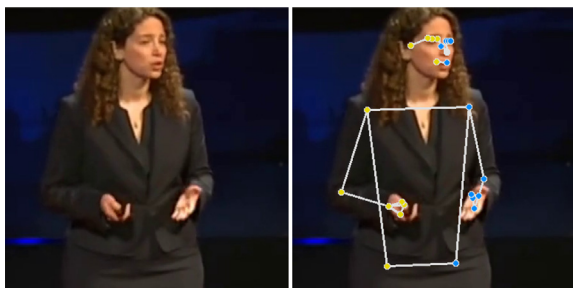


Fig. 2. Example representative poses. Images on the right show the landmarks of the representative frames on the left.



Fig. 3. Sample motion transfer; the left column shows the representative frames from driving (top) and source (bottom) videos; the top right image shows the driving video, and the bottom right image shows the transferred output.

represent. For example, in Case 1, we use E_N (source) and E_H (driving) to produce output A_{E1} and E_N (source) and E_L (driving) to produce output B_{E1} . Outputs A_{E1} and B_{E1} are compared based on their performance in representing high extraversion. We have the output pairs of A_{X_i} and B_{X_i} , where $X \in O, C, E, A, N$ representing personality and $i \in 1, 2, 3, 4, 5$ representing the different cases.

We report ANOVA statistics of our second study in Table 5, using two groupings: The first one separates the samples into different case groups (i) and uses personality (X) as the independent variable; the second one separates the samples into different personality groups and uses different cases as the independent variable. Separating into case groups helps analyze the performance of each approach in expressing distinct factors; for example, we observe that Cases 3 and 4 are the most successful ones in distinguishing different factors; on the other hand, the influence of Case 2 is very similar for each factor. Separating into personality groups helps analyze the expressibility of each personality dimension; for example, we observe that our cases best distinguish extraversion while the smallest influence is on conscientiousness. The linearity and sphericity conditions of repeated measures ANOVA hold for both groupings.

We further inspect pairwise differences with Tukey HSD; Table 6 shows the mean differences of personality factors per case and the mean differences of cases per personality factor. We observe Case 3 having the highest mean differences between different factors. In Case 3, Output A utilizes X_H for the driving input, and Output B uses the

Table 5

ANOVA statistics for the second user study considering case (first five rows) and personality (last five rows) groups.

| | Source | SS | df | F | η^2 |
|---|----------------|--------|-----|----------|----------|
| 1 | Between groups | 7.37 | 4 | 3.305* | 0.083 |
| | Within groups | 80.86 | 145 | – | – |
| 2 | Between groups | 7.13 | 4 | 2.375 | 0.061 |
| | Within groups | 108.86 | 145 | – | – |
| 3 | Between groups | 10.37 | 4 | 10.632** | 0.226 |
| | Within groups | 35.36 | 145 | – | – |
| 4 | Between groups | 25.22 | 4 | 10.662** | 0.227 |
| | Within groups | 85.76 | 145 | – | – |
| 5 | Between groups | 17.10 | 4 | 6.824** | 0.158 |
| | Within groups | 90.86 | 145 | – | – |
| O | Between groups | 20.49 | 4 | 9.019** | 0.199 |
| | Within groups | 82.36 | 145 | – | – |
| C | Between groups | 8.24 | 4 | 3.732** | 0.093 |
| | Within groups | 80.03 | 145 | – | – |
| E | Between groups | 39.16 | 4 | 17.518** | 0.325 |
| | Within groups | 81.03 | 145 | – | – |
| A | Between groups | 21.37 | 4 | 10.167** | 0.219 |
| | Within groups | 76.20 | 145 | – | – |
| N | Between groups | 28.04 | 4 | 12.380** | 0.254 |
| | Within groups | 82.10 | 145 | – | – |

** Statistics marked indicate $p < 0.001$.

* Statistics marked indicate $p < 0.05$.

driving input X_L . For extraversion, A_{E3} is preferred for expressing the high traits, showing the importance of motion for this factor. In the same case, Output B is preferred for agreeableness and neuroticism, suggesting these factors are more related to appearance. In Case 2, where we utilize the same driving motion, agreeableness and neuroticism are easily distinguished due to the high trait source input. Openness and conscientiousness are barely realized in Cases 2 and 3.

In Cases 1, 4, and 5, we expect the means to be close to 0, as in Output A, we utilize the high trait sample as driving input, and the source inputs of both outputs are the same. Extraversion behaves as expected; this is not true for other factors. For example, using low trait source input for conscientiousness and agreeableness factors causes Output B to be slightly confused for high traits, causing their means to get close to 1; however, the difference between Cases 1, 4, and 5 is not significant.

Cases 1 and 2 significantly differ in extraversion and agreeableness; extraversion is better distinguished when the motions differ, and agreeableness is better when appearances differ. The results of the different cases agree that the driving input can express extraversion better, and the source input represents agreeableness and neuroticism better. Cases 1, 2, and 5 are not successful in creating a distinction in conscientiousness; this can be due to this factor being hard to observe in short sequences. Additionally, our original samples had a lower variance in conscientiousness, which can be the reason for this reduced effect. Similarly, openness is not distinguished sufficiently for Cases 2, 3, and 5. Fig. 4 shows box plots for the same groupings.

Extraversion is the most distinguished factor as motion cues represent perceived extraversion well, and all cases except for Case 2 utilize opposing motions; utilizing the same motion, distinguishing extraversion becomes challenging. Agreeableness and neuroticism are better distinguished through appearance cues. Openness is best expressed through motion when neutral or high appearance is used; however, using the low trait sample as the source input diminishes openness perception. Conscientiousness is best observed in Case 3 through its driving input and is hardly distinguished in any other case.

We perform Pearson correlation analysis between the answers to the first and second questions to determine if the output's realism is related to personality perception. Although realism is essential in expressing

Table 6

Pairwise comparisons from the second study and the corresponding ANOVA statistics. We perform ANOVA on different cases where personality factors are independent variables (left half, pairs from {O, C, E, A, N}) and on different personality factors where cases are independent variables (right half, pairs from {1, 2, 3, 4, 5}). The dependent variable is the participants' choices mapped to integers 0, 1, and 2, corresponding to choosing Output A, Equal, and Output B, respectively. We report each pair's mean difference and Tukey HSD adjusted p values. Bold values show $p < 0.05$.

| Grp. | Case 1 | | Case 2 | | Case 3 | | Case 4 | | Case 5 | | Grp. | O | | C | | E | | A | | N | |
|------|-------------|-------------|--------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------|-------|-------------|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | M. | Adj. Diff. | M. | Adj. Diff. | M. | Adj. Diff. | M. | Adj. Diff. | M. | Adj. Diff. | | M. | Adj. Diff. | M. | Adj. Diff. | M. | Adj. Diff. | M. | Adj. Diff. | Mean Diff. | Adj. P |
| C-O | 0.46 | .121 | 0.01 | .999 | 0.30 | .529 | 0.23 | .724 | 0.40 | .243 | 2-1 | 0.40 | .236 | -0.06 | .998 | 0.56 | .001 | -0.86 | .001 | -0.30 | .585 |
| E-O | -0.36 | .330 | -0.20 | .835 | -0.83 | .001 | -0.33 | .388 | -0.76 | .001 | 3-1 | 0.43 | .168 | 0.26 | .756 | -0.03 | .999 | 0.33 | .450 | 0.66 | .011 |
| A-O | 0.70 | .004 | -0.56 | .029 | 0.60 | .018 | 0.36 | .291 | 0.33 | .427 | 4-1 | -0.13 | .958 | -0.36 | .475 | -0.10 | .934 | -0.46 | .135 | 0.43 | .217 |
| N-O | 0.23 | .751 | -0.46 | .112 | 0.46 | .116 | 0.80 | .001 | 0.30 | .535 | 5-1 | 0.23 | .745 | 0.16 | .945 | -0.16 | .687 | -0.13 | .962 | 0.30 | .585 |
| E-C | -0.83 | .001 | -0.20 | .835 | -1.13 | .001 | -0.56 | .024 | -1.16 | .001 | 3-2 | 0.03 | .999 | 0.33 | .570 | -0.60 | .001 | 1.20 | .001 | 0.96 | .001 |
| A-C | 0.23 | .751 | -0.56 | .029 | 0.30 | .529 | 0.13 | .953 | -0.06 | .997 | 4-2 | -0.53 | .049 | -0.30 | .666 | -0.66 | .001 | 0.40 | .264 | 0.73 | .004 |
| N-C | -0.23 | .751 | -0.46 | .112 | 0.16 | .909 | 0.56 | .024 | -0.10 | .985 | 5-2 | -0.16 | .909 | 0.23 | .834 | -0.73 | .001 | 0.73 | .002 | 0.60 | .031 |
| A-E | 1.06 | .001 | -0.36 | .315 | 1.43 | .001 | 0.70 | .002 | 1.10 | .001 | 4-3 | -0.56 | .030 | -0.63 | .041 | -0.06 | .984 | -0.80 | .001 | -0.23 | .784 |
| N-E | 0.60 | .020 | -0.26 | .635 | 1.30 | .001 | 1.13 | .001 | 1.06 | .001 | 5-3 | -0.20 | .837 | -0.10 | .991 | -0.13 | .833 | -0.46 | .135 | -0.36 | .381 |
| N-A | -0.46 | .121 | 0.10 | .985 | -0.13 | .958 | 0.43 | .146 | -0.03 | .999 | 5-4 | 0.36 | .321 | 0.53 | .125 | -0.06 | .984 | 0.33 | .450 | -0.13 | .966 |

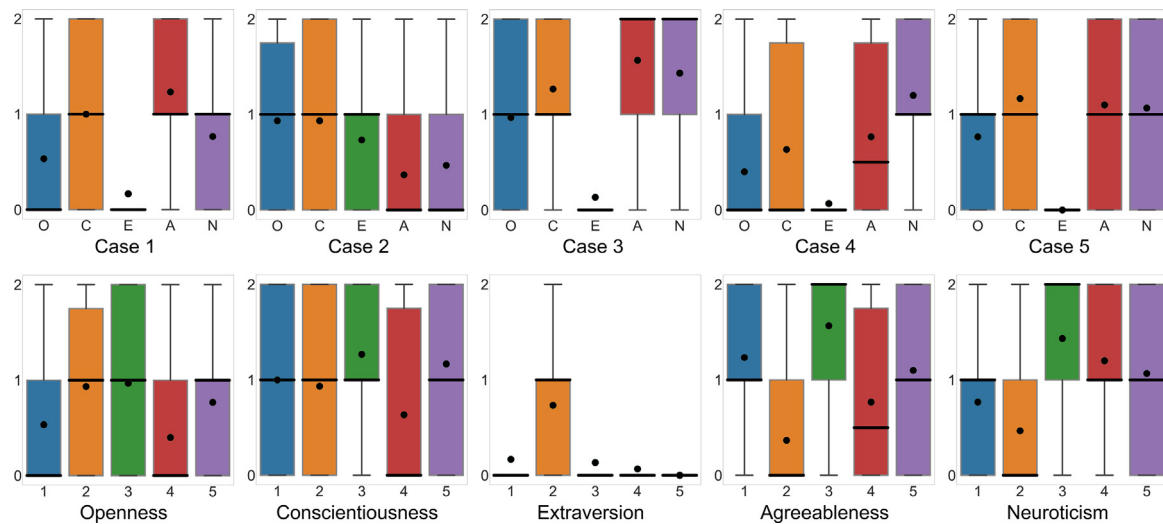


Fig. 4. Box plots for the second study where participants' choice for Output A, Equal, and Output B are mapped to integers 0, 1, and 2, respectively. Black lines depict the median, and the black dots show the mean. Means close to 1 show that the corresponding combination does not help distinguish the high and low traits of the related personality. The first row examines the factors grouped by case, and the second row examines the cases grouped by personality.

personality [89], we do not find a consistent relationship between the participants' personality and realism choices. This observation is likely because, in our case, realism is only harmed by TPS's artifacts when the source and driving inputs are too different in terms of their pose. Unlike a difference in rendering quality or style, the artifacts do not alter the outputs dramatically; the figures' movement and appearance are still readable, with occasional deformations and blending problems. People can read human motion accurately even with minimal information such as point light data [90]. As long as the hands and elbows of the figure are visible, which is the case in our outputs, we do not expect confusion for the motion. Similarly, the deformations do not cause a change in the stylistic elements of the character; thus, we do not expect them to alter the character's image in the participants' minds.

Similar to our personality analysis, we map participants' answers to the integers 0 (Output A), 1 (Equal), and 2 (Output B) for the question asking which output looks realistic compared to others. We report the Pearson correlation coefficients between the participants' answers to realism and personality choices in Table 7. A linear relationship is primarily visible in Case 3, where opposing trait samples are used as driving and source inputs. The correlation is weak for conscientiousness and extraversion and moderate for agreeableness and neuroticism, suggesting some participants may have perceived the personality of the output concerning their realism. On the other hand, such a relationship is not visible in other cases. Especially in Cases 4 and 5, we see a clear preference for the outputs of matching trait inputs having high realism;

Table 7

Pearson correlation coefficients (Cor.) between the participants' answers to realism and personality questions and their P-values. Bold values show $p < 0.05$. We assume $Cor. < 0.4$ as weak and $0.4 < Cor. < 0.8$ as moderate.

| Factor | Case: | 1 | 2 | 3 | 4 | 5 |
|--------|-------|-------------|-------|-------------|-------------|-------|
| O | Cor. | .534 | .145 | -.093 | .417 | .187 |
| | P | .002 | .444 | .625 | .022 | .321 |
| C | Cor. | .338 | .323 | .386 | .436 | .270 |
| | P | .068 | .082 | .035 | .016 | .149 |
| E | Cor. | .193 | .296 | .378 | .240 | .122 |
| | P | .306 | .112 | .040 | .202 | .522 |
| A | Cor. | .339 | .340 | .502 | .315 | .075 |
| | P | .067 | .066 | .005 | .090 | .695 |
| N | Cor. | .055 | -.119 | -.515 | .066 | -.170 |
| | P | .772 | .533 | .004 | .727 | .369 |

Case 4 offers the optimal combination for Output A (High-High), and Case 5 offers the optimal combination for Output B (Low-Low); these outputs received a relatively high preference for realism, yet they do not have a significant correlation to the answers to the personality question. Consequently, we believe the influence of the occasional artifacts in outputs does not have a visible effect on the experiments.

We illustrate some of the outputs that contain occasional artifacts in Fig. 5. We show the consecutive frames of one-second-long output

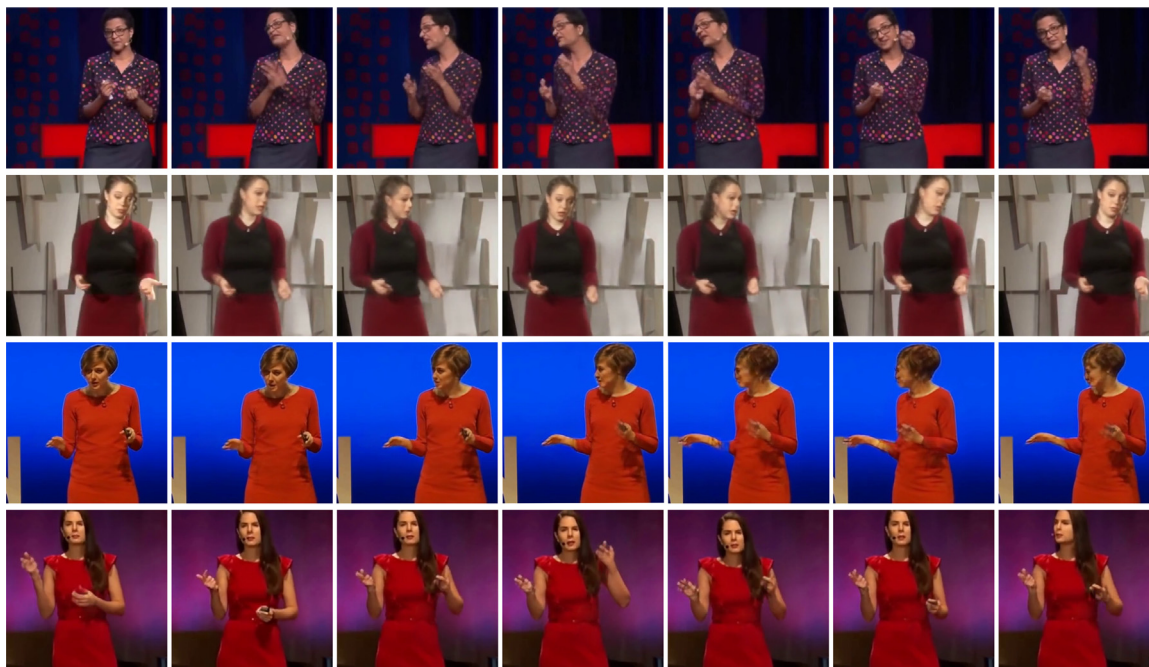


Fig. 5. Consecutive frames of example outputs from TPS. Rows illustrate the frames that cover short segments of the outputs A_{C2} , A_{C1} , A_{O2} , B_{A1} , in order. The first output has artifacts where limb parts are mixed with the background, and the second output has artifacts in the background; the third output has a deformed face due to differences in facing directions. In contrast, the fourth output has no obvious artifacts.

segments that contain artifacts. In the first output, the forearm of the figure is mixed with the cloth texture; this is likely due to TPS's confusion regarding the boundary of cloth and skin. The second output has a non-stationary background due to the figure changing position. The model fills in the background details, but the output can contain artifacts when the original background has high complexity; this is not prominent when the source input has a simple background. The third output contains artifacts due to a mismatch between the facing directions of the source and driving inputs; the source is facing forward with a slightly tilted head, and a part of the driving input requires the head to turn left. The same output also has a problem with the forearm of the figure blending into the background. In contrast, the last output has no obvious artifacts as its source contains a simple background, and its source and driving inputs have well-matching poses. Additionally, the driving input does not have dramatic pose changes, which also helps the output to have better quality. Generally, the artifacts are due to TPS having less information about the source input; future work can improve output quality by supporting multiple inputs from different views to fill in the missing details.

4. Discussion

Our first study approves that participants' zero-acquaintance personality judgments generally agree. Despite minimal information about the individual, participants observe personality traits in short silent videos showing the subject's upper body. In this case, the participants watch the video until the end; the subject's movements, clothes, facial expressions, or the scene's background and lighting can all influence the perceived personality. Future work can dissect each element to analyze their influence individually; however, in this work, we assumed that apparent personality is mainly due to movements and looks.

Our second study analyzes TPS's performance in personality expression using the selected samples. Following our first research question, Cases 1 and 2 results show that TPS can alter personality perception by utilizing different driving and source inputs on neutral appearance and motion, respectively. Participants perceive the videos whose appearance or motion is altered by TPS differently; this suggests that

applications can utilize TPS to alter personality perception of pre-recorded videos. For example, an intelligent assistant expressing different personality traits based on the user may utilize TPS to create variation with less effort.

The output pairs we form using TPS focus on analyzing the influence of two main elements for personality expression: Appearance and motion. Case 1 analyzes the influence of motion by using the opposing driving inputs on the same source input. We observe that the perceived extraversion of the sample highly depends on its motion; the driving input also influences the perceived openness but with less confidence. Case 2 analyzes the influence of appearance using the opposing source inputs on the same driving input. The results suggest that perceived agreeableness and neuroticism highly depend on the subject's appearance. Following our second research question, the impact of motion and appearance on perceived personality is not the same; motion better reflects extraversion and openness, while appearance better reflects agreeableness and neuroticism. This phenomenon is best observed in our Case 3, where the motion and appearance of the output utilize opposing trait samples. The high extraversion motion with the low extraversion appearance is perceived as high extraversion, indicating the importance of motion for extraversion. On the other hand, the opposite is true for agreeableness; high agreeableness appearance with low agreeableness motion is perceived as high agreeableness. Cases 4 and 5 agree with the results of Cases 1 and 2.

We analyze the results of our second user study by grouping the samples according to case and personality. The case groups help us observe whether using different combinations influences the personality factors similarly. We observe that certain cases, such as 3 and 4, cause a higher variance between groups; they are more successful in representing distinct factors. On the other hand, Case 2 has less variance between groups, suggesting the participants perceived the different personality factors together when utilizing the appearance. Following our third research question, the influence of TPS varies for different factors; different cases control different factors dominantly, and the correlation between factors changes based on the case. Grouping by personality helps see the overall effect of TPS on personality perception; a significant difference can be observed for all personality factors using

appearance or motion: the highest distinction is for extraversion, and the lowest is for conscientiousness.

TPS's performance may depend on how much the inputs express the corresponding traits. For example, transferring the motion of a slightly high extraversion sample to the same neutral source can be less effective than transferring the motion of a relatively higher extraversion sample. Future studies can utilize TPS inputs on a wider scale. Although TPS's output had occasional artifacts in certain cases, we did not observe a significant effect due to the realism or quality of the compared outputs. Advances in motion transfer models can enable future work to express certain factors better. For example, the model we used cannot transfer the mimics of the driving input in great detail; however, including such subtle details can improve the performance. Recent work in motion generation utilizes diffusion models to generate text-driven human animations [91–93], and style-based motion [94,95]. Future work can utilize motion generation models to control the stimuli more.

5. Limitations

Although we aim to analyze the implications of personality in altered videos that adopt certain looks and motions with a thorough study, the vastness of the field leads us to use certain simplifications. First, the typical usage of TIPI asks one question for each polarity of personality that is averaged with the opposite trait to measure the related factor more precisely. In our case, the sensitivity and psychometric properties of the instrument are harmed to shorten the time required for evaluating each sample. The average time to complete the first user study for a participant was 37 min.

The feedback from the participants indicated that they found the study long. Since the second user study involved comparisons between output pairs, we resorted to further simplifications. Instead of asking personality questions involving all five factors, we asked a single question focusing on the factor output pairs use. Since we aimed to capture the changes in one personality dimension in each case, this simplification did not harm our analysis; however, with more data captured, we could also analyze if the changes affected other factors simultaneously.

In the second study, we used the choices left, equal, and right instead of using a Likert scale for the participants' answers. While this causes a loss of precision, using a scale for a question that asks which sample expresses the high traits better could have caused confusion. Furthermore, we are interested in whether different input combinations of TPS work similarly for different personality factors rather than measuring how effective each is; consequently, this brief approach helps shorten the average study time without obstructing our analysis. The average time to complete the second study was 18 min.

Our participant count is relatively low for a perception study; we tried to keep the sample size low to compensate for this, which could cause the results not to generalize well in a bigger dataset. In particular, all the individuals used in the videos are TED presenters, which may cause them to have a bias in certain factors like conscientiousness and agreeableness. On the other hand, since we utilize short clips from the talks, the current motion and appearance become more influential on the apparent personality. We also compensate for the low participation count by recruiting participants from reliable sources. Although we ran the study online, all the participants were current university students and their family members. With the spread of paid online services for running user studies, the quality of the responses becomes questionable and requires higher participant counts and further analysis. In contrast, our preference was for quality over quantity.

With a higher participant count, we would utilize non-overlapping groups for the two user studies. While we do not expect a significant difference as the samples used in the first study are not directly used in the second study, there could have been some bias due to familiarizing the individuals briefly. We utilize the samples from the first study as

source inputs in the second study, which could cause those participating in both studies to choose differently. On the other hand, we have a 15-day break between the two studies, which can avoid the familiarity bias for the second study. ANOVA results assuming two groups of participants as the independent variables reveal no significant differences between the answers of those participating in both studies and those participating only in the second study. The ANOVA statistics comparing the two groups of users based on each output pair are available in our repository.

We should also note that the personality ratings gathered in the first study only reflect the apparent personalities of the people in short video segments and are not expected to depict the real personalities of the individuals. As our aim is on the apparent personality in short first encounters, we do not investigate whether these apparent cues reflect the self-reported personalities or expert assessments; instead, we investigate if motion transfer can alter personality perception for the average user. One of our future interests is to use such a system to express different personalities in video entertainment and education; for example, what happens if we alter the appearance or motion of a tutor teaching a subject?

6. Conclusion

We analyzed the influence of different appearance and motion combinations on the perceived personality. To this end, we carried out two user studies. In the first study, we analyzed the apparent personality traits of a set of short video sequences. Then, we selected the samples that best represent each personality factor's high, neutral, and low traits. Using these samples as a motion transfer network's source and driving inputs, we generated stimuli for our second study. The second study compared TPS's output pairs following different combinations to discover their effect on perceived personality.

Our results suggest that using different combinations for TPS's inputs influences extraversion to the most significant extent; utilizing the driving input helps express this factor in general. When the source and driving inputs utilize opposing trait samples in Case 3, we see that the driving input helps convey high extraversion, and the source input helps express high agreeableness and neuroticism. Using TPS has little influence on perceived openness and conscientiousness. Expressing openness in short sequences is challenging, which can explain the lack of influence on openness. Furthermore, our samples in the first study generally scored high in conscientiousness. They had slight variance since they all included presenters, which could have a diminishing effect on comparing the conscientiousness pairs.

Our results can inspire future work aiming for data-driven personality expression; the motion and appearance of videos can be transformed to express the desired traits, enabling control over nonverbal communication. Motion transfer can enhance informative videos to provide a better learning experience; movie sequences can be improved for better character portrayal. Intelligent assistants can employ motion transfer to express different personalities that help communicate with different user types. Future work can automatically generate personality-specific human motions or appearances, which can benefit from an accurate personality prediction system for labeling. Consequently, designers can adjust a few sliders to generate suitable characters or motions to express the desired personality traits. Note that personality perception depends on many aspects of communication; therefore, a complete analysis of the elements involved in personality expression requires more studies. This work will be one more step towards our goal of better-represented digital personalities.

CRedit authorship contribution statement

Ayda Yurtoğlu: Writing – review & editing, Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Sinan Sonlu:** Writing – review & editing, Writing – original draft, Software,

Methodology, Data curation, Conceptualization. **Yalım Doğan**: Writing – review & editing, Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Uğur Güdükbay**: Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

Data availability

The data used in this research is publicly available as TED TALKS dataset in <https://paperswithcode.com/dataset/ted-talks>.

Acknowledgments

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

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.cag.2024.01.013>.

References

- [1] McCrae RR, Costa PT. Personality in adulthood: A five-factor theory perspective. New York, NY, USA: Guilford Press; 2005.
- [2] Durupinar F, Kapadia M, Deutsch S, Neff M, Badler NI. PERFORM: Perceptual approach for adding OCEAN personality to human motion using Laban Movement Analysis. *ACM Trans Graph* 2016;36(1):16, Article no. 6.
- [3] Smith HJ, Neff M. Understanding the impact of animated gesture performance on personality perceptions. *ACM Trans Graph* 2017;36(4):12, Article no. 59.
- [4] Sonlu S, Güdükbay U, Durupinar F. A conversational agent framework with multi-modal personality expression. *ACM Trans Graph* 2021;40(1):17, Article no. 7.
- [5] Vinciarelli A, Mohammadi G. A survey of personality computing. *IEEE Trans Affect Comput* 2014;5(3):273–91.
- [6] Naumann LP, Vazire S, Rentfrow PJ, Gosling SD. Personality judgments based on physical appearance. *Pers Soc Psychol Bull* 2009;35(12):1661–71.
- [7] Kim HJ, Neff M, Lee SH. The perceptual consistency and association of the LMA effort elements. *ACM Trans Appl Percept* 2022;19(1):17. <http://dx.doi.org/10.1145/3473041>, Article no. 1.
- [8] Erkoç Z, Demirci S, Sonlu S, Güdükbay U. Skeleton-based personality recognition using Laban Movement Analysis. In: *Understanding Social Behavior in Dyadic and Small Group Interactions*, Proceedings of Machine Learning Research, vol. 173, Maastricht, Netherlands: PMLR; 2022, p. 74–87.
- [9] Lawson ED. Hair color, personality, and the observer. *Psychol Rep* 1971;28(1):311–22.
- [10] Pancer SM, Meindl JR. Length of hair and beardedness as determinants of personality impressions. *Percept Mot Skills* 1978;46(3 suppl):1328–30.
- [11] Jones AL, Kramer RS, Ward R. Signals of personality and health: the contributions of facial shape, skin texture, and viewing angle. *J Exp Psychol Hum Percept Perform* 2012;38(6):1353.
- [12] Hu Y, Parde CJ, Hill MQ, Mahmood N, O'Toole AJ. First impressions of personality traits from body shapes. *Psychol Sci* 2018;29(12):1969–83.
- [13] Namatame H, Saito M, Sawamiya Y. Personality traits associated with body shape. *Int J Affect Eng* 2016;15(2):161–6.
- [14] Feinberg RA, Mataro L, Burroughs WJ. Clothing and social identity. *Cloth Text Res J* 1992;11(1):18–23.
- [15] Paunonen SV, Ewan K, Earthy J, Lefave S, Goldberg H. Facial features as personality cues. *J Pers* 1999;67(3):555–83.
- [16] Sha T, Zhang W, Shen T, Li Z, Mei T. Deep person generation: A survey from the perspective of face, pose and cloth synthesis. *ACM Comput Surv* 2021;55(12):37, Article no. 257.
- [17] Kammoun A, Slama R, Tabia H, Ouni T, Abid M. Generative adversarial networks for face generation: A survey. *ACM Comput Surv* 2022;55(5):37, Article no. 94.
- [18] Tewari A, Thies J, Mildenhall B, Srinivasan P, Treitsch E, Yifan W, et al. Advances in neural rendering. *Comput Graph Forum* 2022;41(2):703–35.
- [19] Eshraghian JK. Human ownership of artificial creativity. *Nat Mach Intell* 2020;2(3):157–60.
- [20] Ye Z, Wu H, Jia J. Human motion modeling with deep learning: A survey. *AI Open* 2022;3:35–9.
- [21] He Y, Pereira A, Kucherenko T. Evaluating data-driven co-speech gestures of embodied conversational agents through real-time interaction. In: *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*. New York, NY: ACM; 2022, p. 1–8.
- [22] Toshpulatov M, Lee W, Lee S. Generative adversarial networks and their application to 3D face generation: A survey. *Image Vis Comput* 2021;108:104119.
- [23] Siarohin A, Woodford OJ, Ren J, Chai M, Tulyakov S. Motion representations for articulated animation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE; 2021, p. 13653–62.
- [24] Gosling SD, Rentfrow PJ, Swann Jr WB. A very brief measure of the Big-Five personality domains. *J Res Personal* 2003;37(6):504–28.
- [25] Zhao J, Zhang H. Thin-plate spline motion model for image animation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA, USA: IEEE Computer Society; 2022, p. 3647–56. <http://dx.doi.org/10.1109/CVPR52688.2022.00364>, URL: <https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.00364>.
- [26] Boyle GJ, Matthews G, Saklofske DH. Personality theories and models: An overview. In: *Personality Theory and Assessment*. Personality Theories and Models. vol. 1, Washington, DC, USA: American Psychological Association; 2008, p. 1–29.
- [27] Davidson RJ. Toward a biology of personality and emotion. *Ann New York Acad Sci* 2001;935(1):191–207.
- [28] Kathleen Bishop S, Dixon MR, Moore JW, Lundy MP. Behavioral perspectives on personality. In: *Encyclopedia of Personality and Individual Differences*. Springer; 2020, p. 421–31.
- [29] Bornstein RF, Denckla CA, Chung WJ. Psychodynamic models of personality, In: *Handbook of Psychology*, 2nd ed. vol. 5, Wiley Online Library; 2012.
- [30] Zuckerman M, Kuhlman DM, Joireman J, Teta P, Kraft M. A comparison of three structural models for personality: the big three, the big five, and the alternative five. *J Personal Soc Psychol* 1993;65(4):757.
- [31] Agnew R. The effect of appearance on personality and behavior: Are the beautiful really good? *Youth Soc* 1984;15(3):285–303.
- [32] Funder DC, Colvin CR. Friends and strangers: acquaintanceship, agreement, and the accuracy of personality judgment. *J Personal Soc Psychol* 1988;55(1):149.
- [33] Ready RE, Clark LA, Watson D, Westerhouse K. Self-and peer-reported personality: Agreement, trait ratability, and the “self-based heuristic”. *J Res Personal* 2000;34(2):208–24.
- [34] Pozzebon JA, Ashton MC. Personality and values as predictors of self-and peer-reported behavior. *J Individ Differ* 2009;30(3):122–9.
- [35] Kandler C, Riemann R, Spinath FM, Angleitner A. Sources of variance in personality facets: A multiple-rater twin study of self-peer, peer-peer, and self-self (dis) agreement. *J Pers* 2010;78(5):1565–94.
- [36] Funder DC. Accurate personality judgment. *Curr Dir Psychol Sci* 2012;21(3):177–82.
- [37] Zhou MX, Mark G, Li J, Yang H. Trusting virtual agents: The effect of personality. *ACM Trans Interact Intell Syst* 2019;9(2–3):1–36.
- [38] Sajjadi P, Hoffmann L, Cimiano P, Kopp S. A personality-based emotional model for embodied conversational agents: Effects on perceived social presence and game experience of users. *Entertain Comput* 2019;32:100313.
- [39] Zibrek K, Kokkinara E, McDonnell R. The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role? *IEEE Trans Vis Comput Graphics* 2018;24(4):1681–90.
- [40] Ferstl Y, Neff M, McDonnell R. ExpressGesture: Expressive gesture generation from speech through database matching. *Comput Animat Virtual Worlds* 2021;32(3–4):e2016.
- [41] Ghorbani S, Ferstl Y, Carboneau MA. Exemplar-based stylized gesture generation from speech: An entry to the GENE Challenge 2022. In: *Proceedings of the International Conference on Multimodal Interaction*. New York, NY: ACM; 2022, p. 778–83.
- [42] Milcent AS, Kadri A, Richir S. Using facial expressiveness of a virtual agent to induce empathy in users. *Int J Hum-Comput Interact* 2022;38(3):240–52.
- [43] Thakkar JH, Rao PSB, Shubham K, Jain V, Jayagopi DB. Understanding interviewees' perceptions and behaviour towards verbally and non-verbally expressive virtual interviewing agents. In: *Companion Publication of the 2022 International Conference on Multimodal Interaction*. New York, NY: ACM; 2022, p. 61–9.
- [44] Gloor PA, Fronzetti Colladon A, Altuntas E, Cetinkaya C, Kaiser MF, Ripberger L, et al. Your face mirrors your deepest beliefs—Predicting personality and morals through facial emotion recognition. *Future Internet* 2022;14(1):5.
- [45] Hyde J, Carter EJ, Kiesler S, Hodgins JK. Evaluating animated characters: Facial motion magnitude influences personality perceptions. *ACM Trans Appl Percept* 2016;13(2):17. <http://dx.doi.org/10.1145/2851499>, Article no. 8.
- [46] Suen HY, Hung KE, Lin CL. TensorFlow-based automatic personality recognition used in asynchronous video interviews. *IEEE Access* 2019;7:61018–23.
- [47] Hu Y, O'Toole AJ. First impressions: Integrating faces and bodies in personality trait perception. *Cognition* 2023;231:105309.
- [48] Giraud T, Focone F, Demulier V, Martin JC, Isableu B. Perception of emotion and personality through full-body movement qualities: A sport coach case study. *ACM Trans Appl Percept* 2015;13(1):27. <http://dx.doi.org/10.1145/2791294>, Article no. 2.

- [49] McDonnell R, Jörg S, Hodgins JK, Newell F, O'sullivan C. Evaluating the effect of motion and body shape on the perceived sex of virtual characters. *ACM Trans Appl Percept* 2009;5(4):14. <http://dx.doi.org/10.1145/1462048.1462051>, Article no. 20.
- [50] Zibrek K, Niay B, Olivier AH, Hoyet L, Pettre J, McDonnell R. The effect of gender and attractiveness of motion on proximity in virtual reality. *ACM Trans Appl Percept* 2020;17(4):15. <http://dx.doi.org/10.1145/3419985>, Article no. 14.
- [51] Koppensteiner M. Motion cues that make an impression: Predicting perceived personality by minimal motion information. *J Exp Soc Psychol* 2013;49(6):1137–43.
- [52] Neff M, Wang Y, Abbott R, Walker M. Evaluating the effect of gesture and language on personality perception in conversational agents. In: *Proceedings of the 10th International Conference on Intelligent Virtual Agents*. Berlin, Germany: Springer; 2010, p. 222–35.
- [53] Wang Y, Tree JEF, Walker M, Neff M. Assessing the impact of hand motion on virtual character personality. *ACM Trans Appl Percept* 2016;13(2):23. <http://dx.doi.org/10.1145/2874357>, Article no. 9.
- [54] Hachimura K, Takashina K, Yoshimura M. Analysis and evaluation of dancing movement based on LMA. In: *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*. Piscataway, NJ: IEEE; 2005, p. 294–9.
- [55] Aristidou A, Charalambous P, Chrysanthou Y. Emotion analysis and classification: understanding the performers' emotions using the LMA entities. *Comput Graph Forum* 2015;34(6):262–76. <http://dx.doi.org/10.1111/cgf.12598>.
- [56] Dewan S, Agarwal S, Singh N. Spatio-temporal laban features for dance style recognition. In: *Proceedings of the 24th International Conference on Pattern Recognition*. Piscataway, NJ: IEEE; 2018, p. 2911–6.
- [57] Ajili I, Mallem M, Didier JY. Human motions and emotions recognition inspired by LMA qualities. *Vis Comput* 2019;35(10):1411–26.
- [58] Thomas S, Ferstl Y, McDonnell R, Ennis C. 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*. Piscataway, NJ: IEEE; 2022, p. 11–20.
- [59] Lang Y, Liang W, Wang Y, Yu LF. 3D face synthesis driven by personality impression. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33. Washington, DC: AAAI Publications; 2019, p. 1707–14.
- [60] Durupinar F. Personality-driven gaze animation with conditional generative adversarial networks. 2020, arXiv preprint [arXiv:2012.02224](https://arxiv.org/abs/2012.02224).
- [61] Calsius F, Asteriadis S. Synthesizing personality-dependent body postures using generative adversarial networks. In: *Proceedings of the 31st Benelux Conference on Artificial Intelligence and the 28th Belgian Dutch Conference on Machine Learning*. CEUR workshop proceedings, Aachen, Germany: CEUR; 2019, p. 1–15.
- [62] Zabala U, Rodriguez I, Martínez-Otzeta JM, Lazkano E. Expressing robot personality through talking body language. *Appl Sci* 2021;11(10):4639.
- [63] Perez Garcia M, Saffon Lopez S. Exploring the uncanny valley theory in the constructs of a virtual assistant personality. In: *Intelligent Systems and Applications: Proceedings of the Intelligent Systems Conference*. vol. 1, Berlin, Germany: Springer; 2020, p. 1017–33.
- [64] Mori M, MacDorman KF, Kageki N. The uncanny valley [from the field]. *IEEE Robot Autom Mag* 2012;19(2):98–100.
- [65] Seymour M, Yuan L, Dennis AR, Riemer K. Crossing the uncanny valley? Understanding affinity, trustworthiness, and preference for more realistic virtual humans in immersive environments. In: *Proceedings of the 52th Hawaii International Conference on System Sciences*. Atlanta, GA: Association for Information Systems; 2019, p. 1748–58.
- [66] Higgins D, Zhan Y, Cowan BR, McDonnell R. Investigating the effect of visual realism on empathic responses to emotionally expressive virtual humans. In: *Proceedings of the ACM Symposium on Applied Perception*. 2023, p. 1–7.
- [67] Paetzel-Prüsmann M, Perugia G, Castellano G. The influence of robot personality on the development of uncanny feelings. *Comput Hum Behav* 2021;120:106756.
- [68] Amadou N, Haque KI, Yumak Z. Effect of appearance and animation realism on the perception of emotionally expressive virtual humans. In: *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. New York, NY, USA: ACM; 2023, p. 8, Article no. 9.
- [69] Deng Y, Yang J, Chen D, Wen F, Tong X. Disentangled and controllable face image generation via 3D imitative-contrastive learning. 2020, [arXiv:2004.11660](https://arxiv.org/abs/2004.11660).
- [70] Geng Z, Cao C, Tulyakov S. 3D guided fine-grained face manipulation. 2019, [arXiv:1902.08900](https://arxiv.org/abs/1902.08900).
- [71] Thies J, Zollhöfer M, Stamminger M, Theobalt C, Nießner M. Face2Face: Real-time face capture and reenactment of RGB videos. 2020, [arXiv:2007.14808](https://arxiv.org/abs/2007.14808).
- [72] Doukas MC, Zafeiriou S, Sharmanska V. HeadGAN: One-shot neural head synthesis and editing. 2021, [arXiv:2012.08261](https://arxiv.org/abs/2012.08261).
- [73] Liu L, Xu W, Zollhoefer M, Kim H, Bernard F, Habermann M, et al. Neural rendering and reenactment of human actor videos. 2019, [arXiv:1809.03658](https://arxiv.org/abs/1809.03658).
- [74] Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. 2018, [arXiv:1711.09020](https://arxiv.org/abs/1711.09020).
- [75] Nirkin Y, Keller Y, Hassner T. FSGAN: Subject agnostic face swapping and reenactment. 2019, [arXiv:1908.05932](https://arxiv.org/abs/1908.05932).
- [76] Ha S, Kersner M, Kim B, Seo S, Kim D, MarioNETte: Few-shot face reenactment preserving identity of unseen targets. 2019, [arXiv:1911.08139](https://arxiv.org/abs/1911.08139).
- [77] Zakharov E, Shysheya A, Burkov E, Lempitsky V. Few-shot adversarial learning of realistic neural talking head models. 2019, [arXiv:1905.08233](https://arxiv.org/abs/1905.08233).
- [78] Chan C, Ginosar S, Zhou T, Efros AA. Everybody dance now. 2019, [arXiv:1808.07371](https://arxiv.org/abs/1808.07371).
- [79] Ren Y, Yu X, Chen J, Li TH, Li G. Deep image spatial transformation for person image generation. 2020, [arXiv:2003.00696](https://arxiv.org/abs/2003.00696).
- [80] Zhu Z, Huang T, Shi B, Yu M, Wang B, Bai X. Progressive pose attention transfer for person image generation. 2019, [arXiv:1904.03349](https://arxiv.org/abs/1904.03349).
- [81] Siarohin A, Lathuilière S, Tulyakov S, Ricci E, Sebe N. Animating arbitrary objects via deep motion transfer. 2019, [arXiv:1812.08861](https://arxiv.org/abs/1812.08861).
- [82] Siarohin A, Lathuilière S, Tulyakov S, Ricci E, Sebe N. First order motion model for image animation. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché Buc F, Fox E, Garnett R, editors. *Advances in Neural Information Processing Systems*, vol. 32, Red Hook, NY: Curran Associates, Inc.; 2019, p. 7137–47.
- [83] Bansal A, Ma S, Ramanan D, Sheikh Y. Recycle-GAN: Unsupervised video retargeting. 2018, [arXiv:1808.05174](https://arxiv.org/abs/1808.05174).
- [84] Tulyakov S, Liu MY, Yang X, Kautz J. MoCoGAN: Decomposing motion and content for video generation. 2017, [arXiv:1707.04993](https://arxiv.org/abs/1707.04993).
- [85] Yang Z, Zhu W, Wu W, Qian C, Zhou Q, Zhou B, et al. TransMoMo: Invariance-driven unsupervised video motion retargeting. 2020, [arXiv:2003.14401](https://arxiv.org/abs/2003.14401).
- [86] Ni H, Liu Y, Huang SX, Xue Y. Cross-identity video motion retargeting with joint transformation and synthesis. 2022, [arXiv:2210.01559](https://arxiv.org/abs/2210.01559).
- [87] Tukey JW. Comparing individual means in the analysis of variance. *Biometrics* 1949;5(2):99–114.
- [88] Google Developers. MediaPipe. 2023, URL: <https://developers.google.com/mediapipe>.
- [89] Zibrek K, Martin S, McDonnell R. Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Trans Appl Percept* 2019;16(3):19, Article no. 14.
- [90] Blake R, Shiffrar M. Perception of human motion. *Annu Rev Psychol* 2007;58:47–73.
- [91] Zhang M, Cai Z, Pan L, Hong F, Guo X, Yang L, et al. MotionDiffuse: Text-driven human motion generation with diffusion model. 2022, arXiv preprint [arXiv:2208.15001](https://arxiv.org/abs/2208.15001).
- [92] Yuan Y, Song J, Iqbal U, Vahdat A, Kautz J. PhysDiff: Physics-guided human motion diffusion model. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, p. 15964–75.
- [93] Tevet G, Raab S, Gordon B, Shafir Y, Cohen-or D, Bermano AH. Human motion diffusion model. In: *Proceedings of the Eleventh International Conference on Learning Representations*. 2023, URL: <https://openreview.net/forum?id=SJ1kSyO2jwu>.
- [94] Ghorbani S, Ferstl Y, Holden D, Troje NF, Carbonneau MA. ZeroEGGS: Zero-shot example-based gesture generation from speech. *Comput Graph Forum* 2023;42(1):206–16. <http://dx.doi.org/10.1111/cgf.14734>.
- [95] Ao T, Zhang Z, Liu L. GestureDiffuCLIP: Gesture Diffusion Model with CLIP Latents. *ACM Trans Graph* 2023;42(4):18. <http://dx.doi.org/10.1145/3592097>, Article no. 42.