Understanding the Impact of Animated Gesture Performance on Personality Perceptions

HARRISON JESSE SMITH, University of California, Davis
MICHAEL NEFF, University of California, Davis

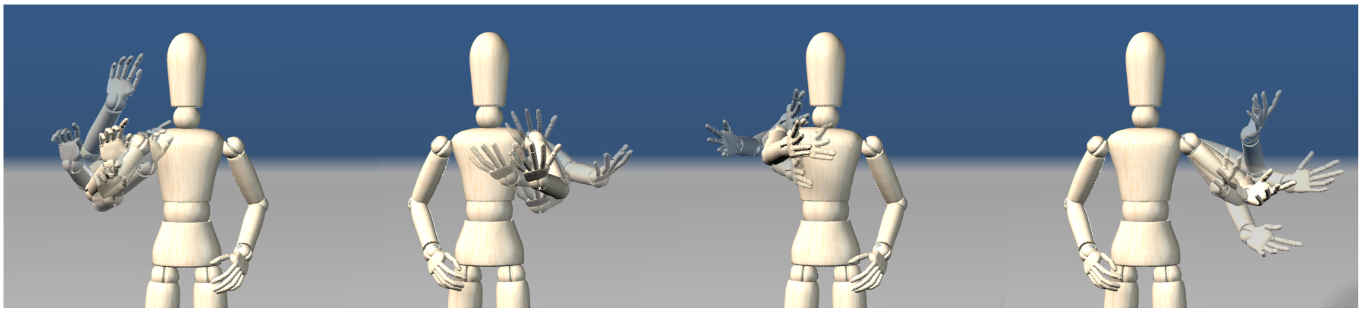

Fig. 1. Gesture performance influences personality perception. We encode changes in gesture performance into animation stimuli and, through perceptual studies, identify which motion adjustments are most relevant to personality. Composited above are four gestures from an unaltered performance and four spatially-warped variations of each used within our study.

Applications such as virtual tutors, games, and natural interfaces increasingly require animated characters to take on social roles while interacting with humans. The effectiveness of these applications depends on our ability to control the social presence of characters, including their personality. Understanding how movement impacts the perception of personality allows us to generate characters more capable of fulfilling this social role. The two studies described herein focus on gesture as a key component of social communication and examine how a set of gesture edits, similar to the types of changes that occur during motion warping, impact the perceived personality of the character. Surprisingly, when based on thin-slice gesture data, people's judgments of character personality mainly fall in a 2D subspace rather than independently impacting the full set of traits in the standard Big Five model of personality. These two dimensions are *plasticity*, which includes extraversion and openness, and *stability*, which includes emotional stability, agreeableness, and conscientiousness. A set of motion properties is experimentally determined that impacts each of these two traits. We show that when these properties are systematically edited in new gesture sequences, we can independently influence the character's perceived stability and plasticity (and the corresponding Big Five traits), to generate distinctive personalities. We identify motion adjustments salient to each judgment and, in a series of perceptual studies, repeatedly generate four distinctly perceived personalities. The effects extend to novel gesture sequences and character meshes, and even largely persist in the presence of accompanying speech. This paper furthers our understanding of how gesture can be used to control the perception of personality and suggests both the potential and possible limits of motion editing approaches.

CCS Concepts: • **Computing methodologies → Animation**; *Perception*;

DOI: http://dx.doi.org/10.1145/3072959.3073697

Additional Key Words and Phrases: Gesture performance, perceptual study, OCEAN personality, crowdsourcing

ACM Reference format:
Harrison Jesse Smith and Michael Neff. 2017. Understanding the Impact of Animated Gesture Performance on Personality Perceptions. *ACM Trans. Graph.* 36, 4, Article 49 (July 2017), 12 pages.
DOI: http://dx.doi.org/10.1145/3072959.3073697

1 INTRODUCTION

At its heart, character animation should be about creating great *characters* - rich, nuanced, and convincing personalities that can make a lasting impression on an audience and lead them to reflect on the human condition. Supporting this goal is a c hallenging research problem, largely because our understanding of how to use movement to convey personality is limited. Artists may generate rich characters for particular scenes, but generating interactive characters that remain consistent across novel situations requires computational support. In this work, we use perceptual studies to understand how changes in animated movement impact the perception of character personality.

Conversational gestures are studied as they play a central role in nonverbal communication [Knapp et al. 2013] and hence are a key factor in social settings where personality is likely to be conveyed. A carefully selected sample of gesture forms was chosen to span the main dimensions of conversational gesture: beats, deictics, metaphorics and iconics [McNeill 1992]. A set of motion edits was then chosen that typifies the adjustments made during animation techniques like motion warping and style transfer. Through a series of perceptual studies, we explored the impact of this set of editing operations on the perception of character personality. We use the Big Five (or OCEAN) model to characterize personality [Costa and McCrae 1992; Wiggins 1996], which is the dominant model of personality used in psychology research and has been the focus of extensive study. The model consists of five factors: extraversion, emotional stability (or neuroticism reversed), agreeableness, conscientiousness, and openness to experience. Adjectives corresponding to each of these traits are shown in Table 1 [Goldberg 1990].

This work provides three key findings. First, these motion manipulations impact the perceived personality of characters along a 2D subspace of the Big Five model of personality, rather than the full 5D space. The subspace dimensions correspond to *stability*, consisting largely of emotional stability, agreeableness, and conscientiousness, and *plasticity*, consisting of extraversion and openness to experience. This suggests both the potential and limitations of these types of motion adjustments and techniques that rely on them, such as motion warping: they can impact the 2D personality space of *stability* and *plasticity*, but the majority of their impact is restricted to this space. This has strong implications for the types of manipulations that must be made to characters to span the full range of personality (e.g. needing changes in clothing, vocal style, etc.) and the role of movement alone in capturing style. Second, we identify which motion adjustments impact the two dimensions of stability and plasticity and how. We show that we can reliably impact the perception of personality along these two dimensions for novel animation sequences by manipulating these factors. These studies were performed on a wooden mannequin to obscure potentially misleading cues from the model. Our third result shows that the movement manipulations largely carry over to more realistic character meshes and animations that include body movement and audio. These results suggest that gesture plays a key role in character personality perception and suggest how to effectively change gesture style to manipulate perceived personality.

Table 1. Adjectives associated with the Big Five traits.

Trait	Low Level	High Level
Extraversion	Reserved, Passive	Assertive, Active
Openness	Shallow, Ignorant	Creative, Cultured
Conscientiousness	Inefficient, Careless	Controlled, Careful
Emotional Stability	Neurotic, Anxious	Calm, Peaceful
Agreeableness	Critical, Malicious	Friendly, Helpful

2 BACKGROUND

Gesture is a type of nonverbal communication employed during discourse to aid communicative intent [Kendon 2004]. However, like many types of nonverbal communication, gesture form and usage is influenced by much more than the conscious intentions of the performer [Ekman and Friesen 1969]. If gesture merely reflected the desire to visually communicate, interlocutors with no visual connection would not gesture at all; yet blind dyads and telephone users are regularly observed employing gesture [de Ruiter 1995; Pérez-Pereira and Conti-Ramsden 2013]. One possibility is that gesture, like all forms of nonverbal communication, is an evolutionarily-derived mechanism by which the internal state of the performer is externally expressed for the benefit of others [Darwin et al. 1998]. Indeed, emotions, attitudes, intentions, and personality have all been shown to subconsciously influence nonverbal behavior [Ekman 1992; Ekman and Friesen 1969; Mehrabian 1968].

Observers can decode these internal states with surprising speed and precision. *Thin-slice* psychology, a sub-field of personality psychology, explores the rate and accuracy with which observers perceive internal states based on very limited information. The stimuli used often take the form of short videos, audio recordings, or photographs. These judgments can be surprisingly accurate [Borkenau and Liebler 1993] and occur very quickly- sometimes in less than 50ms [Borkenau et al. 2009].

The animation community has extended these studies to investigate perceptions of virtual agents: a reasonable extension considering the degree to which artificial stimuli are often subconsciously perceived as real [Reeves and Nass 1996]. A considerable body of work has focused on the perceptual impacts of various character modalities, including body shape (e.g.[McDonnell et al. 2009]), facial expression (e.g. [Pelachaud et al. 1996]), model type [Hodgins et al. 1998], presence of anomalies [Hodgins et al. 2010], spatio-temporal correspondence, [Gielniak et al. 2013], and walking/dancing style [Hoyet et al. 2013]. Studies specifically focusing on personality perception have investigated facial structure (e.g. [Oosterhof and Todorov 2008]), linguistic style (e.g. [Walker et al. 1997]), finger movements [Wang et al. 2016], kinematic pattern [Giraud et al. 2015], walk cycle [Thoresen et al. 2012], and correlations with parameters of Laban Movement Analysis [Durupinar et al. 2016].

Many researchers employ a decoding approach where they capture movement that displays particular aspects of personality and try to distill what factors in the movement lead to this perception (e.g. [Kiiski et al. 2013; Koppensteiner and Grammer 2010; Thoresen et al. 2012]). Other studies (including this one) take an encoding approach, whereby variations are algorithmically inserted and their effects validated through perceptual studies. Hartmann, Mancini, and Pelachaud [2005] used such an approach to map single dimensions of expressivity onto animation parameters. Other studies have focused on mapping specific modalities to personality perception. Mairesse and Walker [2007] generated spoken sentences based upon a desired personality perception. Thoresen, Vuong, and Atkinsonet [2012] adjusted light-point walk cycles to induce different personality perceptions. Guy et al. [2011] derived mappings between crowd-simulation parameters and personality descriptors corresponding to the Eysenck Three-factor personality model. Xu et al. [2013] conducted a user study in which participants adjusted the gesturing style of an NAO robot to express positive, neutral, and negative moods. Neff et al. [2010] adjusted the perception of extraversion using a psychology literature-derived collection of language and movement variations. A similar study adjusted gesture performance to impact emotional stability perceptions [Liu et al. 2015; Neff et al. 2011]. The encoding-based approach employed here allowed us to study the individual impacts of a wide range of motion warping operations on the five primary aspects of personality.

While many personality models exist, the Big Five (or OCEAN) model (Table 1) has emerged as the de facto standard among both personality psychologists and virtual agent researchers [Badler et al. 2002; Costa and McCrae 1992; Wiggins 1996]. This five-factor model is attractive to computer scientists because of its quantitative nature and the orthogonality of the traits. However, two caveats must accompany its use with virtual agents: it was developed to describe the personalities of real people, and the orthogonality of the model describes the extent to which personalities occur within a population. It is not certain that five unique trait judgments can be reliably inferred from thin-slice stimuli. Guy et al. [2011] proposed a two-dimensional factorization of perceived personality in crowds based upon their user study results. Oosterhof and Todorov [2008] found that only two judgments suffice to explain personality perceptions based on facial structure. Giraud et al. [2015] found that two judgments are sufficient to explain personality perceptions based on coaches performing kinematic patterns. Our work suggests that variation in gesture movement may also lead to personality judgments in only two dimensions. This has strong implications for how movement can be used to convey personality and the limits of what may be achieved with gesture motion warping alone.

In PERFORM, Durupinar et al. [Durupinar et al. 2016] present a procedural system for generating personality-based animation, based on Laban Movement Analysis (LMA). LMA provides a high level description of movement and the PERFORM system models the Effort and Shape components of LMA. Perceptual studies were used to learn a mapping between personality and LMA Effort qualities. These Effort qualities can then be changed in their animation system to vary the perceived personality of the animation. An interesting feature of their approach is that a different low-level animation system could be used if it captured LMA variation and it should also produce personality variation. The LMA model varies a large set of movement factors and their study does not directly state which of these factors impact the perceived personality. A major focus of our work is identifying the specific motion adjustments that impact perceived personality. Our work is particularly focused on gesture and also tests stimuli in the presence of spoken audio. They do not identify the two factor model presented here, but further analysis of their data reveals evidence for it (see Sec. 6).

The motion warping operations included in our study were influenced by several different fields. Performing arts literature provides a great deal of knowledge on how movement can be used to generate specific characters (e.g. [Laban and Ullmann 1971; Neff 2014; Stanislavski 2013; Thomas and Johnston 1981; Williams 2009]), but it generally takes skilled interpretation to turn this knowledge into animation. Psychologists have also studied how movement relates to personality (e.g. [Borkenau and Liebler 1993; Knapp et al. 2013; Koppensteiner and Grammer 2010; Riggio and Friedman 1986]), but this literature is neither comprehensive nor at the level of detail necessary to directly produce animation models. The most commonly studied personality trait is extraversion, which correlates with increased gesture stroke scale, and increased distance of elbows and hands from body, clavicle use, fast movements, and loose walking styles [Knapp et al. 2013; Lippa 1998; Riggio and Friedman 1986]. Disfluency in speech, which is often strongly synchronized with gesture, have been observed in anxious speakers [Cappella and Palmer 1990; Furnham 1990] and anxiety is related to neuroticism. Luck et al. [2010] observed that music-induced movements of neurotic individuals tend to be jerkier. Koppensteiner and Grammer [2010] reported that less active performers exhibiting less vertical arm movement are seen as more agreeable, and figures with more pronounced changes in movement direction are perceived as more open. Borkenau and Liebler [1993] noted that relaxed posture is perceived as less conscientious and more emotionally stable.

While influenced by the literature, the motion warping operations included in this study were ultimately selected to capture as much of the natural variation of gestures as possible. These operations can easily be applied to novel motion sequences and are typical of the underlying motion warping operations used with style transfer: the process by which input motion is transformed into a new style without destroying the original content [Brand and Hertzmann 2000; Hsu et al. 2005]. The results of this study provide important guidance about how such operations can affect personality perception.

3 EXPERIMENTAL FRAMEWORK

3.1 Stimuli Creation

All studies required participants to rate animation clips featuring specific manipulations of gesture performance. Motion capture data, performed by a single male actor with a background in gesture and movement studies, was used as input for the first experiment. The second experiment also incoporated performance data from an additional male and a female actor. Finger data was recorded using a pair of Cyberglove II data gloves, each containing 18 bend sensors. Body motion was captured using a 12 camera Vicon optical system with 41 markers placed according to the Vicon human template. The data was manually post-processed to identify the beginning and end of each gesture stroke. Customized sequences of gestures were constructed from individual strokes by procedurally adding preparation and retraction phases as needed. Resulting sequences were approximately 10 seconds in length, similar to previous works on movement-based personality perception [Durupinar et al. 2016]. Because hand and arm motions are most relevant for gestural expression [McNeill 1992], the captured leg, torso, and head motion was not incorporated into any stimuli except the Shakespeare and Road Runner sequences in Experiment 2.

Excluding the Shakespeare and Road Runner sequences (in which we used all gestures produced by the performer), only neutral gesticulations were included in the sequences. The form of a gesticulation has minimal social regulation and lacks a specific semantic meaning [McNeill 1992]. This makes the fundamental nature of a gesticulation robust to reasonable motion warping operations. Gesticulations that might be perceived as highly valenced, such as a vigorously shaken fist, were excluded to avoid strong emotional associations.

The motion capture data was adjusted procedurally to generate the stimuli. Twelve types of motion adjustment, similar to what can commonly be achieved through motion warping techniques, were selected (Table 2). These motion adjustments were selected because they are easily interpretable, and can provide high-level guidance on how more sophisticated techniques, such as style transfer, can affect personality perception. X, Y, and Z translation were implemented by adjusting the wrist position along the horizontal, vertical, and sagittal directions, respectively. Strokes were scaled relative to their center point. Finger and wrist extension were implemented by altering specific joints angles. Arm swivel was implemented by rotating the arm around the axis between shoulder and wrist. Clavicle lift was implemented by rotating the collarbones up. Velocity warp impacted relative timing by specifying how much of the motion should be complete by 1/4, 2/4, and 3/4 of the stroke duration. Average velocity was adjusted by increasing or decreasing the amount of time allowed for a stroke, and interpolating frames as needed. Disfluency introduced a specified number of 'overshoot' events into the gesture stroke, where the overshoot distance and rate of correction were also specified. Tension was implemented using a dynamical system with a PD controller to track wrist position, controlling the amount of error and oscillation in the tracking [Neff et al. 2017].

Excluding the Shakespeare and Road Runner stimuli, gesture sequences were rendered on a faceless wooden mannequin against a blue background (Figure 1). This mesh was chosen to avoid the confounding effects of physical appearance and facial expression, which were not part of this study. Videos were rendered in Maya using a 50 mm camera lens centered on the torso. Videos rendered on the wooden mannequin did not include audio. All videos will be made publicly available on the author's website and representative examples are included in the accompanying video.

3.2 Data Collection

All experimental data, excepting one verification study, was collected using Amazon Mechanical Turk, a crowdsourcing platform commonly used to collect perceptual and psychological survey data. Before choosing to accept a survey task (i.e. human intelligence task or HIT), the participants (i.e. turkers) were shown a short description of the task, the estimated amount of time required (4-7 minutes), and the monetary reward for completion (between $0.25 and $0.35).

Upon accepting the HIT, turkers were informed that they would be shown a series of videos in which the same sequence of gestures would be performed in different styles. Turkers were then shown a short example video demonstrating the range of style variation they should expect. This was followed by the stimuli videos, which were presented in random order to avoid ordering effects. Accompanying each video was a short survey, the Ten Item Personality Inventory (TIPI), a validated instrument for assessing the Big Five personality traits [Gosling et al. 2003]. It consists of two seven-point Likert scale questions for each personality trait. These were presented sequentially, and there was no way to rewatch previous videos after completing and submitting its TIPI. There was no limit to the number of times a turker could watch a video while filling out its TIPI. Upon completion of all TIPIs, turkers were presented with a short, optional survey collecting basic demographic information and general comments.

Even though the reliability of data collected on Mechanical Turk can be on par with more traditional methods [Vuurens et al. 2011], special measures must be taken to ensure quality. Because HITs were written in English, they were restricted to turkers located within the United States. To exclude turkers with a history of poor work, HITs were only available to turkers with a 95% or greater approval rating on prior work. It is often of further benefit to include filter questions: questions with obviously correct answers designed to exclude malicious or inattentive participants. However, due to the subjective nature of personality perception, it was difficult to subtly include such a question within our HITs. Instead, we opted to filter turkers using the Pearson correlation coefficient between an individual's scores and mean video scores. This technique relies upon the insight that the stimuli should elicit somewhat similar judgments in all viewers, and participants whose patterns of answers dramatically differ from the mean are not honestly completing the task. Such techniques have used previously in crowdsourced linguistic judgment studies [Sprouse 2011].

We chose a small correlation coefficient cutoff (0.15) in order to retain as much honest variations in responses as possible. If there was no variation in a turker's responses (e.g. they answered every question identically for every video), their data was also filtered by this step. On average, 25% of survey data was excluded by this technique; exact numbers are given in Table 2 and Table 5. While Amazon Mechanical Turk is known to have a high percentage of outliers [Vuurens et al. 2011], we verified the validity of this technique by running Experiment 2's Road Runner HIT twice: once using Mechanical Turk and again with non-Mechanical Turk participants who completed the HIT under controlled conditions. The resulting data from these two HITs were extremely similar, and the Pearson correlation coefficient of all non-Mechanical Turk participants was greater than 0.15, suggesting that this filtration method is valid.

4 EXPERIMENT 1: IMPACT OF MOVEMENT FACTORS ON PERSONALITY PERCEPTION

This study explored how the twelve motion adjustments, listed in Table 2, impacted the perception of personality. A sequence of four diverse gestures was selected as stimuli for these initial experiments. Gesture sequences were used instead of single gestures in order to minimize the effect of interactions between a motion adjustment and the form of a specific gesture. While identifying such interactions is useful in its own right, it is outside the scope of the study. For each motion adjustment, two versions of the gesture sequence were rendered with different levels of the adjustment. In order to prevent the gesture from appearing unrecognizable or unnatural, these levels of adjustment were manually selected for each motion adjustment and for each gesture. Each adjustment was investigated with a separate HIT, which presented the two versions of the gesture sequence, along with TIPIs, to the turker. We hypothesized that:

H1- It is possible to significantly affect perceptions of all personality traits using only motion adjustments.

H2- Due to the general orthogonality of the Big Five, it would be possible to independently affect perceptions of each personality trait using only motion adjustments.

4.1 Results

Differences in mean scores between pairs of motion-adjusted gesture sequences are shown in Figure 2. False Discover Rate control [Benjamini and Hochberg 1995] was used to correct for Type 1 error. The number of turkers included in the analysis and p_{adj} values are given in Table 2. Seven turkers participated in two experiments; the remaining 428 turkers participated in one.

Significant effects were observed for perceptions of all five personality traits, confirming our first hypothesis. Extraversion was significantly affected by Y translation, finger extension, stroke scale, Z translation, and average velocity. Openness was significantly affected by stroke scale, tension, disfluency, and wrist extension.

Table 2. The different motion adjustments considered in Experiment 1, along with the amount each gesture was adjusted and the significance of the effects. N reports both the total number of turkers recruited, and the amount that was included in analysis after filtering. E = Extraversion, O = Openness, C = Conscientiousness, ES = Emotional Stability, A = Agreeableness.

Motion Adjustment	**N**	**Adjustment Differences Between Videos**				p_{adj}				
		Gesture 1	**Gesture 2**	**Gesture 3**	**Gesture 4**	**E**	**O**	**C**	**A**	**ES**
X Translation	30 / 40	26cm	33cm	20cm	24cm	0.063	0.687	0.563	0.367	0.453
Y Translation	30 / 39	34cm	16cm	16cm	37cm	**<0.001**	0.325	0.327	0.631	0.327
Z Translation	26 / 40	20cm	17cm	13cm	33cm	**0.049**	0.940	0.674	0.855	0.855
Stroke Scale	33 / 55	200%	140%	140%	200%	**<0.001**	**0.049**	0.664	0.911	0.438
Finger Extension	32 / 40	31°	23°	23°	32°	**<0.001**	0.104	0.356	0.426	0.918
Tension	47 / 57	- original kinematic vs. loosely tracked (see 3.1) -				0.057	**0.015**	0.163	**0.029**	**<0.006**
Arm Swivel	42 / 66	————————— 39° —————————				0.316	0.078	0.561	**0.002**	**0.005**
Disfluency	34 / 44	2,4,4	1,4,8	1,4,8	2,5,8	0.664	0.101	**<0.001**	**<0.001**	**<0.001**
Clavicle Lift	30 / 40	————————— 60° —————————				0.453	0.191	**<0.001**	**<0.001**	**<0.001**
Velocity Warp	35 / 55	——— 0.25, 0.50, 0.75 to 0.30, 0.65, 0.85———				0.125	0.844	0.057	**0.035**	**<0.001**
Average Velocity	47 / 69	———————— 2.5x speedup ————————				**<0.001**	0.664	0.890	**<0.001**	**<0.002**
Wrist Extension	49 / 70	40°	29°	23°	34°	0.367	**0.049**	0.555	0.163	0.127

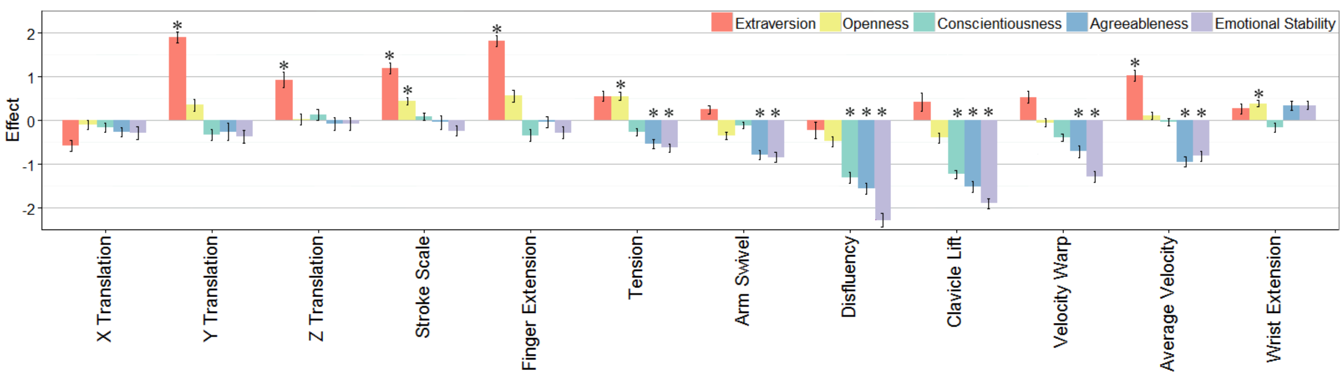

Fig. 2. Effect denotes difference of mean score (on a seven-point scale) between corresponding videos in Experiment 1. Error bars indicate standard error of the mean. Significant effects ($\mathrm{p}_{adj} < 0.05$) denoted by stars.

For both agreeableness and emotional stability, significant changes occurred when adjusting tension, arm swivel, disfluency, clavicle lift, velocity warp, and average velocity. Conscientiousness was significantly affected by disfluency and clavicle lift.

4.2 Correlations in Personality Perception

The data does not definitively address our second hypothesis: that it is possible to independently influence all five personality traits by adjusting gesture performance. As a follow-up analysis, we opted to perform a principal component analysis (PCA) upon the TIPIs collected from all 12 HITs (N=870). Because Big Five traits are generally treated as orthogonal, we anticipated that all five principal components would need to be retained, and that a single personality trait would load primarily onto each.

However, PCA revealed that the first two principal components (PCs) accounted for ~75% of total variance. Retaining only these components was judged sufficient by the Kaiser-Guttman criterion and scree plot inspection (all scree plots are included as a supplemental document), two tests commonly used by behavioral psychologists to determine the number of PCs to retain [Zwick and Velicer 1986]. The PCs were subjected to varimax rotation and the resulting loadings are presented in Table 3. After observing the surprising degree to which the 5D personality scores lay upon a 2D subspace, we performed the same analysis on two additional, distinct data sets to determine whether this substructure was a fluke of the data or indicative of something more general.

The second data set had been previously collected in an early attempt to explore the structure of higher-order interactions between motion adjustments, as well as to test the viability of a between-subjects experimental design. While the stated goals of this experiment were not achieved, the data gathered was still suitable for PCA; the sample size was large (N=840 TIPIs after filtering) and the videos were moderately different from those used in Experiment 1 (different sequence of gestures, a different camera angle, and multiple motion adjustments were randomly applied to each video).

The third data set comprised TIPI scores for a single video. This video contained unadjusted motion capture data, corresponding to the exact hand and arm movements performed by the actor. The purpose of performing PCA on this data set (N=745 after filtering) was to determine if these two PCs, which occurred within data sets rating many different videos, also occurred for ratings of a single video. Stated differently, we wanted to determine if these two PCs describe the structure of within-video variance as well as between-video variance.

The two additional data sets also contained very significant correlations. The Kaiser-Guttman criterion and scree plot inspection indicated that two PCs were sufficient in both cases (explaining 77% and 76% of the variance, respectively). The PC loadings of these additional data sets are extremely similar to the first and are given in Table 3.

Table 3. Similar principal components explained ~ 75% of variance in all three data sets analyzed (Section 4.2). E = Extraversion, O = Openness, C = Conscientiousness, ES = Emotional Stability, A = Agreeableness.

	Experiment 1		Data Set 2		Data Set 3	
	PC1	**PC2**	**PC1**	**PC2**	**PC1**	**PC2**
E	-0.15	0.87	-0.14	0.83	-0.16	0.88
O	0.20	0.49	0.14	0.54	0.23	0.46
C	0.38		0.33		0.37	
ES	0.64		0.66		0.65	
A	0.62		0.63		0.60	

4.3 Conclusion

Multiple gesture motion modifications were shown to significantly affect the perception of each of the Big Five traits. As expected, extraversion was most affected by adjustments which increased the amount of effort/energy required (spreading fingers, increasing velocity/stroke size, moving the gesture upwards or towards the listener).

However, not all of the literature-based adjustments affected extraversion as expected. Outward arm swivel and clavicle use had very little effect. Of particular interest, increasing the width of a gesture (X translation) had a negative (though insignificant) effect on perceived extraversion. This runs counter to much of the extraversion literature, which states that expansive movements are more extraverted [Lippa 1998; Luck et al. 2010; Riggio and Friedman 1986]. It should be noted that the gestures employed were performed one-handed, and the effect of width may be more relevant in the performance of two-handed gestures, where it may more clearly increase the size of the subject's interaction region. Inversely, while there is not much literature suggesting that straightened fingers are correlated with extraversion, it was one of the most potent motion adjustments investigated.

Disfluency and large clavicle lift had the most significant effects overall, achieving high levels of significance for conscientiousness, emotional stability, and agreeableness. Turkers described videos employing clavicle lift as 'less calm', and described videos including disfluencies as 'more hesitant' and 'less confident'. Outward arm swivel decreased agreeableness and emotional stability. This was expected for agreeableness, as previous research has linked the arms-akimbo stance to disagreeable interactions [Mehrabian 1968].

The effect of average velocity on agreeableness and emotional stability was not expected. However, this may be an artifact caused by the abrupt juxtaposition of fast gesture strokes and slow, controlled transitions between gestures. One turker referred to these 'abrupt' gestures as more 'argumentative and unstable'. The negative effects of loose tension on emotional stability and agreeableness seem sensible: a less controlled motion may be viewed as less stable and expressing less concern towards others.

It is perhaps more difficult to explain the alterations affecting openness. The larger level of wrist extension resulted in much more camera-directed palm visibility in the final video, which could have made the agent seem more open. Stroke scale and tension have significant effects upon openness, perhaps because both lead to larger and looser motions that may appear more encompassing and less controlled and hence more open.

One of the most important results of the analysis was determining that the majority of the collected personality ratings lie upon a 2D manifold. Additional analysis found this same pattern in two additional, distinct data sets. The presence of this 2D subspace suggests that, in the context of brief, information-limited virtual agents encounters, people mainly make two distinct judgments of personality. The first judgment principally influences emotional stability, agreeableness, and conscientiousness perceptions, and the second extraversion and openness perceptions.

Similar higher-order factors within the Big Five has been reported in personality psychology literature [Digman 1997], and their existence is believed to be a consequence of genetics and neurobiological substrates [DeYoung 2006; Jang et al. 2006]. The first factor is often referred to as *stability* and is interpreted as a person's emotional (emotional stability), social (agreeableness), and motivational (conscientiousness) stability. The second factor is often referred to as *plasticity* and is interpreted as a person's tendency to explore or voluntarily engage with novel situations . The virtual humans used in the studies have no genetics or biology, but previous experience may have conditioned participants to expect traits to occur in specific combinations.

A more perceptually-focused explanation is the 'Halo Effect' [Thorndike 1920], in which the 'general merit' of a person is used as a heuristic for more specific qualifications or abilities when relevant cues are absent. With this in mind, the second PC may express perceptions of extraversion, arguably the trait with the most visible cues in this setting. Given their natural correlation, extraversion may be acting as a heuristic for openness, explaining its high loading on the second PC. The first PC would then correspond to the perceived overall 'general merit' of the virtual agent. Whatever the true reason behind their existence, the presence of these perceptual covariations is important and useful. They suggest the extent to which different personalities can be created using gesture performance alone.

Taken together, these results show that A) specific changes in gesture performance lead to specific changes in perceived character personality and B) the types of motion adjustments explored here - which underly most motion warping techniques - mainly impact the perception of two distinct personality judgments: stability and plasticity.

5 EXPERIMENT 2: CRAFTING DISTINCT PERSONALITIES

Table 4. Motion adjustments utilized for high and low levels of the stability and plasticity constellations. The mean and standard deviation of the magnitude of the motion adjustments upon all 36 gestures used within the 6 different sequences are also presented.

High Stability	**Low Stability**
No Clavicle Lift	High Clavicle Lift (M=28.8°, SD=3.9°)
Smooth Strokes	Jerky Strokes (M=2.7, SD=0.8; M=4.5, SD=2.0; M=6.1, SD=1.6)
Inward Arm Swivel (M=-9.0°, SD=0°)	Outward Arm Swivel (M=10.7°, SD=2.4°)
High Plasticity	**Low Plasticity**
Large Strokes (M=114%, SD=3%)	Small Strokes (M=85%, SD=0%)
Outward Gestures (M=6cm, SD=3cm)	Inward Gestures (M=-8cm, SD=4cm)
Raised Gestures (M=15cm, SD=7cm)	Lowered Gestures (M=-12cm, SD=8cm)
Extended Fingers (M=10.3°, SD=4°)	Curled Fingers (M=-10.8°, SD= 4°)

We would like to be able to generate characters that reflect particular personalities. If a viewer's personality perceptions reflect two distinct underlying judgments, it should be possible to create four uniquely perceived personalities corresponding to crossed high and low levels of each judgment. We therefore selected two constellations of motion adjustments, intended to independently influence each judgment, based on the motion adjustments effects that proved significant in our first study.

Borrowing terms from the psychology realm, we refer to the first constellation as the stability factor: it includes clavicle lift, disfluency, and wide arm swivel and was intended to influence the perception of emotional stability, agreeableness, and conscientiousness. We refer to the second constellation as the plasticity factor: it includes stroke scale, Z translation, Y translation, and finger extension and was intended to influence extraversion and openness perceptions. The principle behind these selections was to include the most significant motion adjustments that did not affect traits of both judgments. Average velocity, for example, was excluded because it had strong effects on extraversion in addition to emotional stability and agreeableness.

Text descriptions of the adjustments, along with the means and standard deviations of adjustment magnitudes, are given in Table 4. As in the previous experiment, the magnitude of the adjustments was manually selected per gesture and per motion adjustment in order to prevent the performance from appearing unnatural. By performing these adjustments, we created a 'high' and 'low' level for each of the two factors. Crossing these factors allowed us to create four distinct performances for the same gesture sequence, which were used as stimuli in the next set of experiments. To view all videos used in this experiment, please refer to the supplemental video.

Our hypotheses concerning these factors were:

H1- By adjusting the plasticity factor, it is possible to affect perceptions of extraversion and openness without affecting perceptions of conscientiousness, agreeableness, and emotional stability.

H2- By adjusting the stability factor it is possible to affect perceptions of concientiousness, agreeableness, and emotional stability without affecting perceptions of extraversion and openness.

H3- These effects will persist even if the virtual agent uses additional modes of communication/expression beyond gesture.

H4- These effects are not an artifact of online crowdsourced data collection and are reproducible using non-crowdsourced participants.

We tested these hypotheses with six different gesture sequences. Each contained unique gestures, with no overlap between the sequences, or with the gestures used in Experiment 1. In total, 34 different gestures were used. For each sequence, four videos were generated corresponding to fully-crossed high and low levels of the two motion adjustment factors. The magnitude of motion adjustments were manually selected for each gesture, though they were similar in most cases.

The first four sequences were meant to test the first two hypotheses. They were rendered on a wooden mannequin, contained only arm and hand motion and did not include audio. We refer to these sequences as WoodenA, WoodenB, WoodenC and WoodenD.

The fifth and sixth sequences tested the third hypothesis. These videos were rendered on realistic meshes and included hand, arm, torso, head, and lower body motion data. The PD tracker for tension control was used with equal settings for all videos, which had the effect of slightly smoothing out the motion. The timing and ordering of these gestures were unaltered from the original performances, resulting in very faithful representations of the original motions. The audio produced by the performers was included in these sequences, though facial animations and lip syncing were absent, and the face was therefore obscured (Figure 4) to prevent this lack of animation from influencing results. Turkers were instructed to listen to the audio while watching the videos, and were not instructed to weigh gestures over speech when providing their answers. One sequence depicted a male performer delivering a Shakespeare monologue (Julius Caesar 1.3.463-68), and the the other sequence depicted a female performer describing the plot of a Road Runnner cartoon. We refer to these sequences as 'Shakespeare' and 'Road Runner', respectively.

The four different performances of a gesture sequence were presented as a HIT on Amazon Mechanical Turk and filtered as in Experiment 1. The gesture sequence, character mesh, and accompanying audio (if applicable) was kept constant for all videos within the HIT. In order to test our final hypothesis, we repeated the Road Runner HIT using participants who were not crowdsourced online. Nine participants (6M-3F) agreed to take the HIT under controlled conditions (in isolation, free of distractions, while wearing over-the-ear headphones). There was no extra filtering processes for this HIT; answers from all 9 participants were included in the analysis. Interestingly, the Pearson's correlation coefficient between individual and mean scores was quite high for all non-Mechanical Turk participants (M=0.61, SD= 0.16). None were below the filter's cutoff value of 0.15.

5.1 Results

Composites of the mean video scores of all 5 personality dimensions are presented in Figure 3. The left figure, which contains mean scores for videos without speech or full-body motion, composites the scores of videos from WoodenA through WoodenD. The right figure composites scores from Shakespeare, Road Runner, and Road Runner Non-MTurk. Significance was determined using 2-way ANOVAs for each sequence. P-values were adjusted for Type 1 error using False Discovery Rate control and are summarized in Table 5. The total number of participants in each HIT, along with the number remaining after filtering, is also given.

5.2 Conclusion

Our attempts to extend the effects of the stability and plasticity movement adjustments to novel gestures without full-body movement or audio were mostly successful, as shown in Figure 3, where the composite factors impact the expected traits. Extraversion was consistently and exclusively affected by the plasticity adjustment, and emotional stability was consistently and exclusively affected by the stability adjustment. Agreeableness was unexpectedly affected by the plasticity adjustment in one test sequence, WoodenC. This effect might have been the result of the plasticity factor fundamentally altering the first gesture in the sequence. The motion adjustment to this two-handed gesture oriented the palms directly at the camera in the high plasticity (and less agreeable) video, and significantly below the camera in the low plasticity video. Openness was unexpectedly influenced by the stability adjustment in WoodenB and WoodenD, and was not affected by the plasticity adjustment in WoodenC or WoodenD.

The pattern of significant effects changed when full-body movement and audio were included. Extraversion and openness still appeared to be influenced by the plasticity adjustment and emotion stability still appeared to be influenced by the stability adjustment. However, agreeableness was only significantly influenced once and conscientiousness was never significantly influenced. The TIPI adjectives for agreeableness are “critical, quarrelsome" and “sympathetic, warm". The TIPI adjectives for conscientiousness are “dependable, self-disciplined" and “disorganized, careless". It may be the case that cues for these traits are most present in the speech channel and that, in its absence, emotional stability is used as a heuristic.

Though the magnitude of certain effects diminished when additional modalities were present, PCA performed on the data revealed similar principal components for all of the seven HITs, and that similar amounts of variance (75%) were explained by them. We have included the principal component loadings, proportions of variance, mean video score graphs and scree plots for all seven HITs in the supplemental documents.

The results shows that the findings from Experiment 1 are useful in understanding how gesture performance can predictably influence personality perceptions. Perceptions of extraversion and emotional stability can be reliably influenced, even when speech and full body motion is present. In the absense of speech and full body motion, conscientiousness and agreeableness can be influenced using the same motion adjustments used to influence emotional stability. In their presence, however, animators may need to consider varying gesture choice, verbal or other non-verbal behaviors not considered here in order to influence these personality perceptions.

Interestingly, openness appeared to still be aligned with extraversion in the presence of full-body movement and speech. The TIPI adjectives for “Openness to Experience" are “open to new experiences, complex" and “conventional, uncreative". While the researchers can only theorize at this point, the presence of audio may provide some grounding about the actual ideas and concepts being expressed by the avatar. If the topic is suitably mundane, using exaggerated strokes may appear creative and unconventional. Without the grounding provided by speech, the listener might imagine a more exotic topic being discussed, for which exaggerated strokes are more appropriate and conventional.

The similar results of the Road Runner and Road Runner Non-Mechanical Turk HITs suggest that using Pearson’s correlation coefficient between individual and mean scores with a cutoff value of 0.15 is a valid method for filtering out unreliable participants. Not only were the patterns of significance similar, but the correlation coefficients were quite high for all of the Non-Mechanical Turk participants (M=0.61, SD= 0.16). None were lower than 0.15, and thus would not have been filtered out.

The small differences in how motion adjustments affected personality perceptions in different HITs suggests that there may be interaction effects between gesture forms and the edits applied to them. While identifying the specific nature or physical characteristics of a gesture that interact with our motion adjustments is beyond the scope of this work, it is a promising direction for future research.

6 DISCUSSION

This paper presented a set of experiments examining the influence of gesture performance on personality perceptions. The adjustments applied in this study can be classified largely as motion warping: the manner of the gesture performance is changed, but other factors, such as the gesture itself, are not. This work provides two major contributions. First, it identifies multiple gesture adjustments that can be used to influence perceptions of each of the Big Five. Second, the study provides strong evidence that gesture performance allows people to make judgments that lie on a 2D subspace of personality. This suggests possible limits to what can be achieved by motion warping; however, the addition of other modalities and adjustments may provide additional cues that could independently influence personality trait perception. For example, modalities conveying preference and identity claims, such as clothing choice, effectively convey information about openness [Naumann et al. 2009] and the addition of scratching and other self-adaptors have been shown to influence perceptions of emotional stability, but not agreeableness [Neff et al. 2011].

The discovery of the two factor model allows us to re-examine earlier studies of personality, revealing more evidence for the leading role of stability and plasticity in personality perception. For example, Wang et al. [2016] found that hand motion has very significant impacts on perceptions of each of the Big Five. Many of the motions considered impacted multiple personality traits, however. For example, spread hand poses provided the highest levels of extraversion and openness, but the lowest levels of agreeableness, conscientiousness and emotional stability, where relaxed poses were

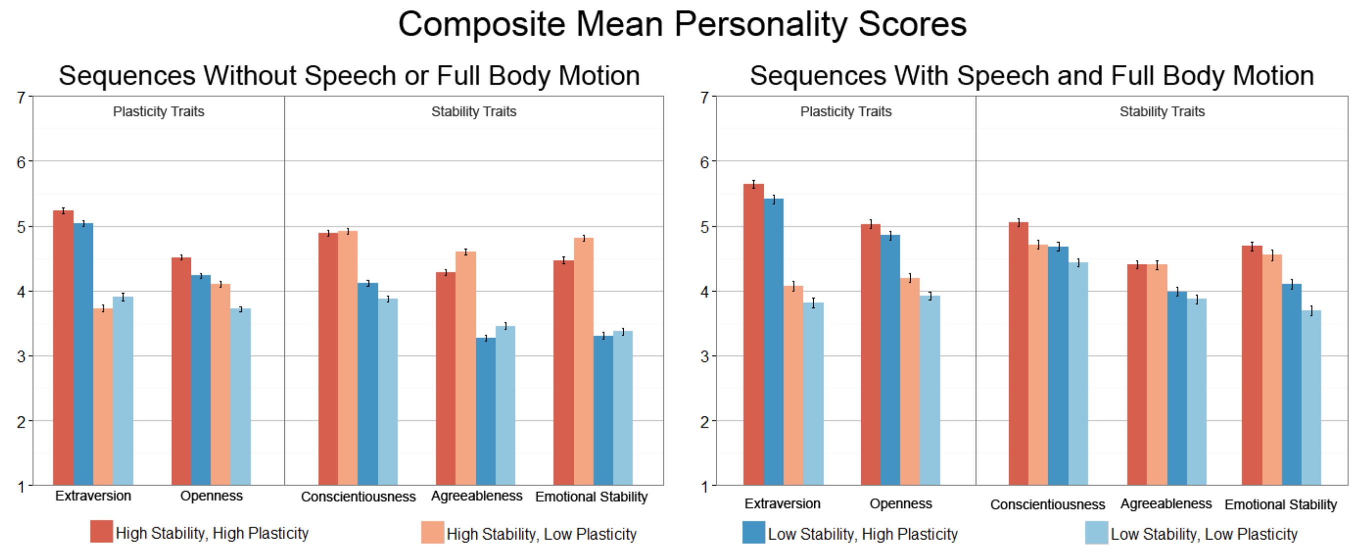

Fig. 3. Mean personality scores for videos used in Experiment 2. Each bar indicates the mean personality score for multiple gesture sequence videos with the same movement adjustments applied to them. The graph on the left composites videos from the WoodenA, WoodenB, WoodenC, and WoodenD sequences. The graph on the right composite videos from the Shakespeare, Road Runner, and Road Runner Non-MTurk sequences. The Plasticity motion adjustment, indicated by bar saturation, was expected to mainly influence extraversion and openness. The Stability motion adjustment, indicated by bar color, was expected to mainly influence conscientiousness, agreeableness, and emotional stability. Note that, in order to highlight these trends, bar order has been adjusted for extraversion and openness. Error bars depict standard error of the mean.

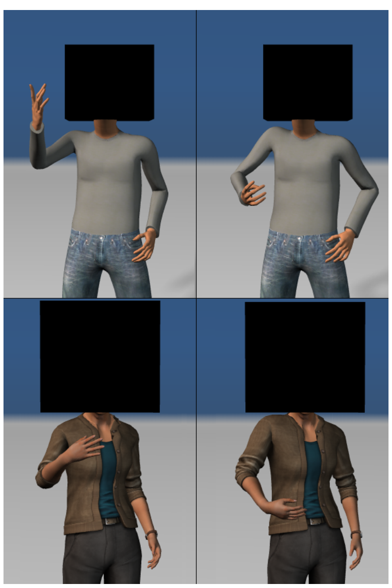

Fig. 4. The same gestures performed with high stability, high plasticity (left) and low stability, low plasticity (right) in the Shakespeare sequence (top) and Road Runner sequence (bottom).

best. Similarly, large motions performed best for plasticity traits and small were preferred for stability traits. A dimensionality reduction analysis of their data may provide further evidence for the 2D substructure reported here.

PERFORM [Durupinar et al. 2016, Fig. 9] did not identify a two factor subspace, but their results seem to suggest that much of the variation observed in their user studies lies in the same two dimensions. Consider the *stability* factors of emotional stability, conscientiousness and agreeableness. Participants were significantly more likely to rate Sustained Time clips as higher in each of these. They were also significantly more likely to rate Direct Space and Bound Flow clips as high in conscientiousness and emotional stability (but not agreeableness, where there was a majority for both factors, but not a significant result). A unique result showed that Light Weight was viewed as more Agreeable. For the *plasticity* factors of Extraversion and Openness, participants rated both higher for clips with Indirect Space and Free Flow, the opposite to the significant results within the stability factors. Participants were more likely to rate Sudden Time clips as extraverted, again, the opposite of the stability factors, but there was no significant effect of Time on openness. Most, but not all, of the variation observed in their study follows the two factor model, revealing a structure that was not previously evident.

7 LIMITATIONS AND FUTURE WORK

There are several limitations and caveats to this work. Although this work used, as its starting point, the five-dimensional OCEAN personality model, there are other models that could have been employed. This work could be reproduced with the three-dimensional PEN model or the six-dimensional HEXACO model by substituting the TIPI with a different survey tool.

The results we have presented only indicate how to adjust perceptions in a relative manner. We restricted ourselves to within-subject

Table 5. Adjusted p-values and F values for the effects of the stability factor, plasticity factor, and their interaction on the different sequences of Experiment 2. N denotes the total number of subjects used in the analysis(after filtering, along with the total number of subjects who participated in the HIT. E = Extraversion, O = Openness, C = Conscientiousness, A = Agreeableness, ES = Emotional Stability.

Sequence (N)	**Movement**	p_{adj} $(F\ value)$				
(F Test Degrees Of Freedom)	**Adjustment**	**E**	**O**	**C**	**A**	**ES**
WoodenA (50 / 57)	Stability	0.20 (2.7)	0.67 (0.3)	<**0.01** (36.9)	<**0.01** (62.1)	<**0.01** (47.8)
(1, 196)	Plasticity	<**0.01** (53.9)	<**0.01** (25.9)	0.20 (2.7)	0.93 (0.0)	0.93 (0.0)
	Interaction	0.25 (2.1)	0.52 (0.7)	0.29 (1.6)	0.67 (0.3)	0.29 (1.6)
WoodenB (29 / 35)	Stability	0.92 (0.0)	<**0.01** (14.6)	<**0.01** (16.8)	<**0.01** (17.9)	<**0.01** (31.3)
(1, 112)	Plasticity	<**0.01** (29.1)	**0.03** (6.5)	0.55 (1.0)	0.39 (1.7)	0.09 (4.4)
	Interaction	0.57 (0.8)	0.63 (0.5)	0.92 (0.0)	0.91 (0.1)	0.55 (0.9)
WoodenC (32 / 35)	Stability	0.61 (0.4)	0.32 (1.8)	**0.01** (9.5)	**0.03** (7.4)	<**0.01** (24.9)
(1, 124)	Plasticity	<**0.01** (49.9)	0.23 (2.6)	0.84 (0.0)	**0.01** (9.8)	0.17 (3.4)
	Interaction	0.61 (0.4)	0.84 (0.0)	0.50 (0.8)	0.62 (0.4)	0.17 (3.2)
WoodenD (34 / 40)	Stability	0.20 (2.9)	<**0.01** (10.7)	<**0.01** (49.5)	<**0.01** (51.7)	<**0.01** (69.6)
(1, 132)	Plasticity	<**0.01** (40.9)	0.09 (4.5)	0.88 (0.0)	0.53 (1.3)	0.74 (0.2)
	Interaction	0.74 (0.3)	0.74 (0.2)	0.74 (0.6)	0.74 (0.4)	0.88 (0.0)
Shakespeare (38 / 42)	Stability	0.97 (0.0)	0.97 (0.0)	0.26 (2.9)	0.06 (6.1)	**0.05** (6.8)
(1, 148)	Plasticity	<**0.01** (90.0)	<**0.01** (18.7)	0.26 (2.7)	0.47 (1.5)	0.78 (0.5)
	Interaction	0.97 (0.0)	0.96 (0.1)	0.60 (1.0)	0.81 (0.4)	0.97 (0.1)
Road Runner (31 / 47)	Stability	0.27 (3.1)	0.26 (2.8)	0.27 (2.5)	0.26 (2.3)	<**0.01** (10.2)
(1, 120)	Plasticity	<**0.01** (34.5)	**0.01** (17.2)	0.41 (1.1)	0.31 (1.6)	0.27 (2.0)
	Interaction	0.59 (0.47)	0.58 (0.6)	0.74 (0.2)	0.94 (0.0)	0.27 (2.2)
Road Runner Non-MTurk (9 / 9)	Stability	0.12 (3.9)	0.06 (5.8)	0.12 (3.3)	**0.02** (8.6)	**0.02** (8.7)
(1, 32)	Plasticity	**0.01** (12.1)	<**0.01** (14.8)	0.06 (5.6)	0.12 (3.3)	0.31 (1.7)
	Interaction	0.34 (1.3)	0.42 (1.0)	1.00 (0.0)	1.00 (0.0)	0.70 (0.3)

experimental designs, as early experiments revealed extremely high variance in the personality scores assigned by different turkers. We expect these experiments would produce similar results using a between-subject design, but would need a much larger sample size to average out variations in how people use Likert scales. What would likely be more difficult is mapping a particular movement change to an absolute personality value due to fluctuations in how people map personality (or any abstract concept) to a numeric scale. By way of comparison, people may readily agree that banana bread taste like bananas, and be able to say which of two breads has a stronger banana flavor. However, it would be difficult to get agreement on whether the bread should have an absolute value of four or five on the seven-point scale of banana-ness.

Second, our stimuli included only low-affect gestures. The adjustment effects described here may not hold for high valence gestures. We observed this in an initial exploratory experiment: a backhand slap, when raised, lowered agreeableness very significantly. Future studies may specifically examine high valence gestures. There may well be a complex relationship between content (the gesture form) and performance.

A final limitation is that there is currently no automatic way to determine how much a gesture can be altered while remaining recognizable. In this study, the amount of the modification applied to each gesture was set manually and tweaked as needed to avoid rendering a gesture unrecognizable or unnatural. A long-term goal of this line of research is the creation of a computational model that can modulate gesture with minimal human input. For example, the meaning of some gestures is changed by small variations of the palm orientation, but this may have no impact on other gestures. To achieve this, we must better understand what features of specific gestures can be varied and which are invariant. This is a non-trivial problem: gestures with similar forms, yet different meanings underlying them, may have different invariants. In addition, this problem is complicated by the sheer number and amount of variation between gestures. Future work could address this issue by using an active learning approach, coupled with a crowdsourced rating system, to build a cleverly parameterized model of the variant and invariant aspects of certain gestures.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Nos. IIS 1115872 and IIS 1320029. We gratefully acknowledge their support. We would like to thank the anonymous reviewers for their thoughtful input, our performers, and our study participants. This work benefited greatly from many useful discussions with Marilyn Walker, Jean E. Fox Tree and Simine Vazire.

REFERENCES

Norman Badler, Jan Allbeck, Liwei Zhao, and Meeran Byun. 2002. Representing and parameterizing agent behaviors. In *Computer Animation, 2002. Proceedings of.* IEEE, 133–143.

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)* (1995), 289–300.

Peter Borkenau, Steffi Brecke, Christine Möttig, and Marko Paelecke. 2009. Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality* 43, 4 (2009), 703–706.

Peter Borkenau and Anette Liebler. 1993. Consensus and Self-Other Agreement for Trait Inferences from Minimal Information. *Journal of Personality* 61, 4 (1993), 477–496.

Matthew Brand and Aaron Hertzmann. 2000. Style machines. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques.* ACM Press/Addison-Wesley Publishing Co., 183–192.

J Cappella and M Palmer. 1990. The structure and organization of verbal and non-verbal behavior: Data for models of production. *Handbook of language and social psychology* (1990), 141–161.

Paul T Costa and Robert R McCrae. 1992. Four ways five factors are basic. *Personality and individual differences* 13, 6 (1992), 653–665.

Charles Darwin, Paul Ekman, and Phillip Prodger. 1998. *The expression of the emotions in man and animals.* Oxford University Press, USA.

Jan de Ruiter. 1995. Why do people gesture at the telephone? *Proceedings of the CLS opening Academic Year 1995-1996* (1995).

Colin G DeYoung. 2006. Higher-order factors of the Big Five in a multi-informant sample. *Journal of personality and social psychology* 91, 6 (2006), 1138.

John M Digman. 1997. Higher-order factors of the Big Five. *Journal of personality and social psychology* 73, 6 (1997), 1246.

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 (TOG)* 36, 1 (2016), 6.

Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.

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

Adrian Furnham. 1990. Language and personality. (1990).

Michael J Gielniak, C Karen Liu, and Andrea L Thomaz. 2013. Generating human-like motion for robots. *The International Journal of Robotics Research* 32, 11 (2013), 1275–1301.

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 (TAP)* 13, 1 (2015), 2.

Lewis R Goldberg. 1990. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology* 59, 6 (1990), 1216.

Samuel D Gosling, Peter J Rentfrow, and William B Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

Stephen J Guy, Sujeong Kim, Ming C Lin, and Dinesh Manocha. 2011. Simulating heterogeneous crowd behaviors using personality trait theory. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation.* ACM, 43–52.

Björn Hartmann, Maurizio Mancini, and Catherine Pelachaud. 2005. Implementing expressive gesture synthesis for embodied conversational agents. In *gesture in human-Computer Interaction and Simulation.* Springer, 188–199.

Jessica Hodgins, Sophie Jörg, Carol O'Sullivan, Sang Il Park, and Moshe Mahler. 2010. The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception (TAP)* 7, 4 (2010), 22.

Jessica K Hodgins, James FO Brien, and Jack Tumblin. 1998. Perception of human motion with different geometric models. *Visualization and Computer Graphics, IEEE Transactions on* 4, 4 (1998), 307–316.

Ludovic Hoyet, Kenneth Ryall, Katja Zibrek, Hwangpil Park, Jehee Lee, Jessica Hodgins, and Carol O'Sullivan. 2013. Evaluating the distinctiveness and attractiveness of human motions on realistic virtual bodies. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 204.

Eugene Hsu, Kari Pulli, and Jovan Popović. 2005. Style translation for human motion. In *ACM Transactions on Graphics (TOG)*, Vol. 24. ACM, 1082–1089.

Kerry L Jang, W John Livesley, Juko Ando, Shinji Yamagata, Atsunobu Suzuki, Alois Angleitner, Fritz Ostendorf, Rainer Riemann, and Frank Spinath. 2006. Behavioral genetics of the higher-order factors of the Big Five. *Personality and individual Differences* 41, 2 (2006), 261–272.

Adam Kendon. 2004. *Gesture: Visible action as utterance.* Cambridge University Press.

Hanni Kiiski, Ludovic Hoyet, Brendan Cullen, Carol O'Sullivan, and Fiona N Newell. 2013. Perception and prediction of social intentions from human body motion. In *Proceedings of the ACM Symposium on Applied Perception.* ACM, 134–134.

Mark Knapp, Judith Hall, and Terrence Horgan. 2013. *Nonverbal communication in human interaction.* Cengage Learning.

Markus Koppensteiner and Karl Grammer. 2010. Motion patterns in political speech and their influence on personality ratings. *Journal of Research in Personality* 44, 3 (2010), 374–379.

Rudolf Laban and Lisa Ullmann. 1971. The mastery of movement. (1971).

Richard Lippa. 1998. The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis. *Journal of Research in Personality* 32, 1 (1998), 80–107.

Kris Liu, Jackson Tolins, Jean Fox Tree, Michael Neff, and Marilyn Walker. 2015. Two Techniques for Assessing Virtual Agent Personality. (2015).

G Luck, S Saarikallio, B Burger, MR Thompson, and P Toiviainen. 2010. Effects of the Big Five and musical genre on music-induced movement. *Journal of Research in Personality* 44, 6 (2010), 714–720.

François Mairesse and Marilyn Walker. 2007. PERSONAGE: Personality generation for dialogue. In *Annual Meeting-Association For Computational Linguistics*, Vol. 45. 496.

Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona N Newell, and Carol O'Sullivan. 2009. Investigating the role of body shape on the perception of emotion. *ACM Transactions on Applied Perception (TAP)* 6, 3 (2009), 14.

David McNeill. 1992. *Hand and mind: What gestures reveal about thought.* University of Chicago Press.

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

Laura P Naumann, Simine Vazire, Peter J Rentfrow, and Samuel D Gosling. 2009. Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin* (2009).

Michael Neff. 2014. Lessons from the arts: what the performing arts literature can teach us about creating expressive character movement. In *Nonverbal Communication in Virtual Worlds.* ETC Press, 123–148.

Michael Neff, Pengcheng Luo, Yingying Wang, and Nicholas Toothman. 2017. *Making an Impression: Flexible Generation of Gesture Form and Body Movement.* Technical Report. University of California - Davis, Department of Computer Science.

Michael Neff, Nicholas Toothman, Robeson Bowmani, Jean E Fox Tree, and Marilyn A Walker. 2011. DonâĂŹt scratch! self-adaptors reflect emotional stability. In *International Workshop on Intelligent Virtual Agents.* Springer Berlin Heidelberg, 398–411.

Michael Neff, Yingying Wang, Rob Abbott, and Marilyn Walker. 2010. Evaluating the effect of gesture and language on personality perception in conversational agents. In *Intelligent Virtual Agents.* Springer, 222–235.

Nikolaas N Oosterhof and Alexander Todorov. 2008. The functional basis of face evaluation. *Proceedings of the National Academy of Sciences* 105, 32 (2008), 11087–11092.

Catherine Pelachaud, Norman I Badler, and Mark Steedman. 1996. Generating facial expressions for speech. *Cognitive science* 20, 1 (1996), 1–46.

Miguel Pérez-Pereira and Gina Conti-Ramsden. 2013. *Language development and social interaction in blind children.* Psychology Press.

Byron Reeves and Clifford Nass. 1996. *How people treat computers, television, and new media like real people and places.* CSLI Publications and Cambridge university press.

Ronald E Riggio and Howard S Friedman. 1986. Impression formation: The role of expressive behavior. *Journal of Personality and Social Psychology* 50, 2 (1986), 421.

Jon Sprouse. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods* 43, 1 (2011), 155–167.

Constantin Stanislavski. 2013. *Building a character.* A&C Black.

Frank Thomas and Ollie Johnston. 1981. *Disney animation: The illusion of life.* Abbeville Press.

John C Thoresen, Quoc C Vuong, and Anthony P Atkinson. 2012. First impressions: Gait cues drive reliable trait judgements. *Cognition* 124, 3 (2012), 261–271.

Edward L Thorndike. 1920. A constant error in psychological ratings. *Journal of applied psychology* 4, 1 (1920), 25–29.

Jeroen Vuurens, Arjen P de Vries, and Carsten Eickhoff. 2011. How much spam can you take? an analysis of crowdsourcing results to increase accuracy. In *Proc. ACM SIGIR Workshop on Crowdsourcing for Information Retrieval (CIRâĂŹ11).* 21–26.

Marilyn A Walker, Janet E Cahn, and Stephen J Whittaker. 1997. Improvising linguistic style: Social and affective bases for agent personality. In *Proceedings of the first*

international conference on Autonomous agents. ACM, 96–105.
Yingying Wang, Jean E. Fox Tree, Marilyn Walker, and Michael Neff. 2016. Assessing the Impact of Hand Motion on Virtual Character Personality. *ACM Trans. Appl. Percept.* 13, 2, Article 9 (March 2016), 23 pages. DOI: https://doi.org/10.1145/2874357
Jerry S Wiggins. 1996. *The five-factor model of personality: Theoretical perspectives.* Guilford Press.
Richard Williams. 2009. *The Animator's Survival Kit: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators.* Macmillan.
Junchao Xu, Joost Broekens, Koen Hindriks, and Mark A Neerincx. 2013. *Mood expression through parameterized functional behavior of robots.* IEEE.
William R Zwick and Wayne F Velicer. 1986. Comparison of five rules for determining the number of components to retain. *Psychological bulletin* 99, 3 (1986), 432.