A Conversational Agent Framework with Multi-modal Personality Expression

SINAN SONLU and UĞUR GÜDÜKBAY, Bilkent University, Turkey
FUNDA DURUPINAR, University of Massachusetts Boston, USA

Fig. 1. We adjust voice, body movement, and facial expressions to simulate personalities of virtual agents. The left half shows an extraverted agent, and the right half shows an introverted agent in a scenario involving interaction with a passport officer where the agent responds to questions about his passport and origin. The portraits show the corresponding facial expressions.

Consistently exhibited personalities are crucial elements of realistic, engaging, and behavior-rich conversational virtual agents. Both nonverbal and verbal cues help convey these agents' unseen psychological states, contributing to our effective communication with them. We introduce a comprehensive framework to design conversational agents that express personality through non-verbal behaviors like body movement and facial expressions, as well as verbal behaviors like dialogue selection and voice transformation. We use the OCEAN personality model, which defines personality as a combination of five orthogonal factors of openness, conscientiousness, extraversion, agreeableness, and neuroticism. The framework combines existing personality expression methods with novel ones such as new algorithms to convey Laban Shape and Effort qualities. We perform Amazon Mechanical Turk studies to analyze how different communication modalities influence our perception of virtual agent personalities and compare their individual and combined effects on each personality dimension. The results indicate that our personality-based modifications are perceived as natural, and each additional modality improves perception accuracy, with the best performance achieved when all the modalities are present. We also report some correlations for the perception of conscientiousness with neuroticism and openness with extraversion.

Authors' addresses: S. Sonlu and U. Güdükbay, Department of Computer Engineering, Bilkent University, Bilkent, Ankara, 06800, Turkey; emails: sinan.sonlu@bilkent.edu.tr, gudukbay@cs.bilkent.edu.tr; F. Durupınar, Department of Computer Science, University of Massachusetts, 100 Morrisey Boulevard, Boston, MA, 02125, USA; email: funda.durupinarbabur@umb.edu.

1 INTRODUCTION

Concepts from human psychology, such as emotions and personality, are widely used in research on intelligent agents for creating realistic characters with a rich set of behaviors and effective communication skills. Especially, endowing virtual agents with personality has been shown useful in building rapport and enhancing user experience in various domains [Gratch et al. 2007]. For instance, a virtual character can be equipped with the most favorable personality to build trust with a job applicant [Zhou et al. 2019], increase patients' willingness to disclose health information [Lucas et al. 2014], help reduce anxiety in individuals with autism and developmental disorders [Burke et al. 2017], take part in a virtual audience to alleviate the fear of public speaking [Batrinca et al. 2013], or support creativity by brainstorming ideas taking user personality into consideration [Buisine et al. 2007].

Creating compelling personalities is a multi-faceted task, and the omission of necessary communication channels can be detrimental to the believability of the virtual agent and the communication of the message. In humans, various modalities such as physical appearance and body posture [Naumann et al. 2009], as well as speech [Polzehl 2015] and dialogue content [Mairesse et al. 2007], influence judgments of others' personalities. Previous research on

virtual characters indicates correlations between different modalities of behavior and the perceived personalities of characters. For instance, there is a link between the facial features of virtual characters and personality inference [Todorov et al. 2008]. Similarly, people assess the personalities of animated characters based on their gestures, speech content [Neff et al. 2010], and movement styles [Durupinar et al. 2017] and react to them as they would to real humans [de Borst and de Gelder 2015].

We introduce a framework to create conversational agents that express personality through a comprehensive model consisting of facial expressions, body movement, verbal style, and voice transformation. In technical terms, we translate the agent's personality features into automated modifiers for three-dimensional (3D) animation and auditory parameters, referring to theories from psychology and social sciences to analyze and parameterize each concept. We then evaluate how much each modality contributes to personality perception to understand effective (virtual) communication strategies. Such a model can be used in video games and animated movies to build immersive experiences, and in developing virtual tutors and assistants to improve approachability (e.g., virtual health-care, support desk) and in social virtual environments to represent the individual better.

We use the OCEAN model [McCrae and John 1992] to describe the personality of an agent. We utilize Laban Movement Analysis (LMA) [Adrian 2008; Maletic 2011] to define body motion controlled by OCEAN factors as a formal parameterization between OCEAN and LMA elements that exist in the literature [Durupinar et al. 2017]. For the connection between speech content and OCEAN personality, we refer to the correlations introduced by Mairesse and Walker [2010], who link a set of linguistic cues [Tausczik and Pennebaker 2010] to OCEAN personality traits. We adopt this theory to handcrafted dialogue text that fits each personality type. To configure the vocal features of speech that influence the perceived personality, we use the mapping introduced by Polzehl [2015].

The contributions of this work are twofold:

- We introduce a comprehensive conversational agent framework that allows the creation of virtual characters that interact with human users through various communication channels. The characters are capable of expressing all the five factors of personality through dialogue, voice, body motion, and facial expressions. To our knowledge, no such system that combines all these modalities of personality expression exists to-date. Our system uses both existing methods shown to be effective in conveying personality and novel algorithms. Specifically, we introduce new methods for the expression of Laban Shape and Effort Qualities in addition to updating emotional facial expressions given OCEAN personality values. We define a mapping between personality traits and Laban Shape Qualities to improve personality expression through body motion and validate this through Amazon Mechanical Turk studies. Additionally, we establish affinities between emotional facial expressions and personality through crowd-sourcing experiments on Amazon Mechanical Turk. We provide the source code of our implementation at https://github.com/sinansonlu/Conversational-Agent-Framework.

- We analyze the effectiveness of different communication modalities and their combinations in the perception of virtual human personalities. We show that additional modalities generally improve personality recognition, with the highest performance achieved when all the modalities are combined. We also analyze the correlations between different OCEAN dimensions when different communication channels are active. The results indicate correlations for the perception of conscientiousness with neuroticism, and openness with extraversion. Our evaluations are based on the data that we collected through Amazon Mechanical Turk studies.

In Section 2, we summarize the background and review related work. We introduce the system framework in Section 3, describe the animation modification process in Section 4, explain our experiments and evaluation in Section 5, and provide a general discussion about experimental findings in Section 6. We conclude with future research directions in Section 7.

2 BACKGROUND AND RELATED WORK

2.1 Background

2.1.1 Communication Model. Communication is an encoded message transfer between two sides. Reception can occur through multiple channels, and decoding is required on the receiver's side to interpret the message. Humans communicate through channels of facial expression, body language, voice, verbal style, and verbal content [Ekman 2019]. Key [1975] groups interpersonal communication as *indirect* and *direct*. Indirect communication includes external factors such as the surrounding environment. Direct communication is internal to the sender and the receiver. It includes anything that belongs to or emerges from them. Key [1975] categorizes direct communication into *verbal* and *nonverbal*. The verbal category covers language-related topics, including grammar and word preferences, which form the message as a sentence. This sentence could be spoken or transferred using a language substitute. Key [1975] classifies the nonverbal category into *paralanguage* and *kinesics*. Paralanguage includes non-speech sounds, vocal features, and intonation. Kinesics include all movements resulting from the muscular and skeletal shift, which is encapsulated by the appearance. She examines the muscular activity in the face in a group called *facial expression*, and the remaining movements in *posture*.

We adapt the communication model of Key [1975] to computer animation. An animated scene consists of an environment and at least one agent. The environment covers indirect communication elements of scene properties, lighting, and camera. The agent covers direct communication, including the 3D model, body movement, facial expression, and speech. We further categorize speech as voice and content. The agent's movement, facial expression, and speech can be used expressively to communicate a message encapsulated by personality.

2.1.2 OCEAN Personality Model. Five-factor Personality Model is a commonly used personality classification framework in psychology [McCrae and Costa 2005]. In this model, the personality of an individual is analyzed in five orthogonal dimensions including Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, that form the acronym

OCEAN. Each dimension is two ended, and the low and high ends are explained with multiple personality attributes that are grouped in one factor. The orthogonal and descriptive nature of this model makes it a popular choice in computer science.

Another popular personality model used in the simulation of virtual crowds and interactive agents is the PEN model of personality [Eysenck and Eysenck 1985], which focuses on three dimensions: *psychoticism*, *extraversion*, and *neuroticism*. As an example, Bera et al. [2017] introduce an automatic classifier for predicting pedestrian personality in crowd videos using the PEN model. Both the PEN and OCEAN personality models are factor analytic—their main difference is the factors they represent. The approaches that we employ, such as Mairesse and Walker [2007], Polzehl [2015], and Durupinar et al. [2017], use the OCEAN model of personality; therefore, we also employ this five-factor model.

2.1.3 Laban Movement Analysis. Laban Movement Analysis (LMA) is a framework used for describing human movement. LMA is used by many researchers in computer graphics and robotics to describe humanoid motion [Bernstein et al. 2015; Burton et al. 2016; Durupinar et al. 2017; Lourens et al. 2010; Zhao and Badler 2001]. LMA includes components to analyze motion in terms of spatial and temporal relations of body parts. We can translate these concepts into computer animation using quantitative descriptors for human motion [Larboulette and Gibet 2015]. As a link between the OCEAN model and LMA, we refer to previous research that connects these concepts together [Durupinar et al. 2017].

2.2 Related Work

2.2.1 Expressive Agents. Expressive conversational behavior is a crucial part of realistic communication in animated virtual agents [van Straalen et al. 2009]. Early research in this area has established models that integrate nonverbal elements into communication. For instance, Cassell et al. [1994] couple speech, intonation, facial expressions, and hand gestures in animated conversation; Allbeck and Badler [2002] model the agent's personality, mood, and affect; Gebhard [2005] introduces ALMA—a layered model of affect; and Pelachaud [2005] utilize Affective Presentation Markup Language for expressive nonverbal behavior.

Expressive verbal and non-verbal behaviors in virtual characters convey a wide range of traits such as friendliness and warmth [Randhavane et al. 2019a], competence [Nguyen et al. 2015], and gender [Vala et al. 2011]; as well as increasing the sense of presence in virtual reality environments [Randhavane et al. 2019b]. Appropriate speech and gaze increase the social presence in multi-party interaction [Yumak and Magnenat-Thalmann 2016]. Rendering style is another factor that influences the perception of friendliness, appeal, and realism of virtual characters [McDonnell et al. 2012; Zell et al. 2015]. However, attention must be paid to virtual character design, as incongruous expressions may lead to the exaggeration of the uncanniness effect [Tinwell et al. 2011].

Castellano et al. [2011] show that people can identify the emotional content of synthesized gestures when a motion's expressive features such as fluidity are mapped from the captured motion to synthesized animation. This suggests that motion style can be separated from motion content, and expressed in new motions. Similarly, the computer graphics community has been interested

in establishing a mapping between high-level motion features and apparent personality or emotions. LMA is a popular choice for analyzing and designing expressive motion [Burton et al. 2016; Durupinar et al. 2017; Masuda and Kato 2010] and generally used for parameterizing high-level motion.

2.2.2 Personality Expression in Virtual Agents. Several works employ personality to guide virtual character behavior such as motion, multi-agent interaction, and speech [Durupınar et al. 2016; Gebhard 2005; Shvo et al. 2019]. We compare our work to others that investigate the influence of different communication elements on apparent personality. Among these, Durupinar et al. [2017] describe links between low-level motion parameters and LMA to express OCEAN personality in virtual characters. They focus on the whole-body motion and manipulate animation keyframes to define LMA parameters in collaboration with LMA experts. They perform a user study to establish an OCEAN-LMA mapping, which we adopt in this work. Different from Durupinar et al. [2017], we use a different set of low-level motion parameters to implement LMA qualities and apply Inverse Kinematics to adjust hand motion further. We also include dialogue, vocal adjustments, and facial expressions to investigate the influence of these modalities on apparent personality.

Similarly, Smith and Neff [2017] focus on gesture performance using a set of motion modification parameters mapped onto OCEAN personality. They analyze each motion parameter's influence on apparent personality in short animated clips. They conclude that people's perception of personality happens in a two-dimensional space when it is based only on gestures. This is also compatible with the Big Two model of personality, which consists of plasticity (openness and extraversion) and stability (conscientiousness, agreeableness, and neuroticism). Our findings also support the existence of these two higher order dimensions. Smith and Neff [2017] include the speech of the motion capture actor in one experiment; however, different from our work, they do not investigate the influence of verbal elements on personality. Besides, they exclude the agent's face, while we use it as an additional vessel of personality.

2.2.3 Data-driven Approaches in Expressive Motion. Data-driven techniques have been commonly used to generate expressive agent behavior. Ball and Breese [2001] diagnose the user's emotions and personality to generate appropriate agent behavior using Bayesian networks. Hartmann et al. [2002] synthesize gesture from augmented conversation transcripts utilizing high-level gesture definitions. Lee and Marsella [2006] extract nonverbal behavior rules from real-life video clips that accompany the dialogue. Recently, Burton et al. [2016] generate emotional motion by automatically quantifying LMA parameters; and Randhavane et al. [2019c] generate expressive walking animation for varying dominance levels.

An effective strategy to generate varying motion styles is style transfer [Aberman et al. 2020; Yümer and Mitra 2016], which modifies the style of an existing motion while preserving its intention. Motion datasets with emotion, age, or movement style labels are available [Xia et al. 2015]; yet, there is still a need for personality-annotated motion datasets of adequate size, whether they are 3D motions or 2D full-body videos. Examples such as Escalante et al.

[2018] employ user videos that focus on the face with limited information on the body. As a result, procedural methods are still popular and offer manageable environments for personality perception experiments. As more data becomes available, replacing procedural methods with data-driven techniques will be straightforward thanks to the usage of LMA as an intermediate motion descriptor between personality expression and low-level motion parameters.

2.2.4 Audio and Verbal Features. Verbal interaction with the agent is possible through natural language processing [Best et al. 2020] and is a crucial aspect of expressive communication. The choice of words in social media messages has shown strong connections to personality [Golbeck et al. 2011]. Linguistic cues are effective predictors of OCEAN traits, especially extraversion [Mairesse et al. 2007]. Text that expresses certain personality types can be generated using rule-based language modifiers in a specific context [Mairesse and Walker 2007]. We adopt the linguistic cues described by Mairesse and Walker [2007] to convey personality in agent dialogue.

Additionally, vocal features carry information about personality; and this knowledge is used to both detect personality and to express it. Machine learning models can successfully predict personality based on speech signals [Gilpin et al. 2018]. Through a user study, Polzehl [2015] labels a speech dataset with OCEAN personality factors and extracts audio descriptors such as intensity, pitch, and loudness to train a Support Vector Machine. Polzehl's findings lay the foundation of our vocal feature adjustments. In addition to personality, audio features can be used to express emotions in procedural speech animation [Charalambous et al. 2019].

Although there exist many models and experiments that explore the influence of different subsets of movement, facial expression, dialogue, and voice on personality, a thorough analysis of a comprehensive system that incorporates all these elements is needed. We present such a framework to understand the influence of the combination of such modalities on personality perception. The framework is component-based and open-source so that researchers can test their own models, animations and methods to expand our knowledge of virtual agent communication.

3 SYSTEM FRAMEWORK

The system consists of two main components: *Scenario Handler* and *Animation Modifier*. The input consists of agent personality and user speech. The agent responds to the user with a personality-driven dialogue utterance and expressive animation.

Scenario Handler (see Figure 2) is a state machine that determines the agent's response to the user's speech. Using Watson API [IBM 2015], it transcribes the speech, finds user intent, selects an appropriate response based on the intent, and vocalizes this response by text-to-speech conversion. We train Watson Assistant by setting entities and intents using multiple examples per scenario. For example, for "Show Passport" intent, we train the system using examples such as "May I take a look at your passport?", "Could you give me your passport, please?", and "Where is your passport?" so that it can recognize similar inputs. Entities are the variable words within training examples of intents. They are trained in the same manner and recognized within the intent.
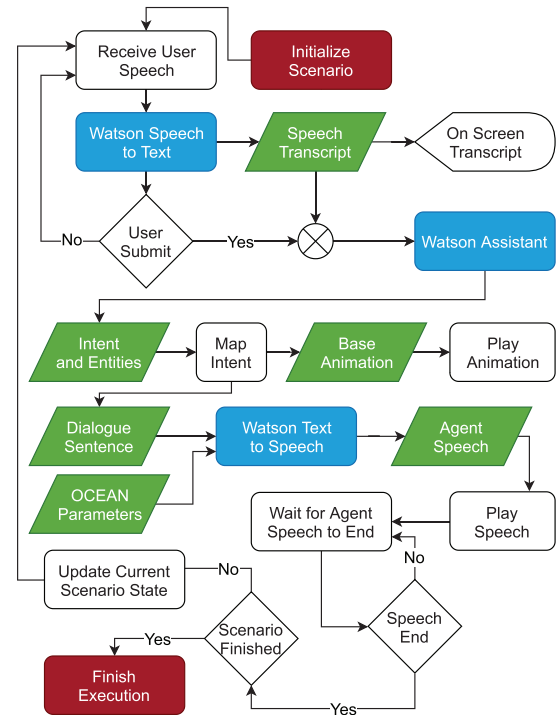

Fig. 2. The Scenario Handler flowchart.

Based on the extracted intent and entities, Scenario Handler determines the corresponding dialogue that is compatible with agent personality and the accompanying base animation without personality-based modifications. Scenario Handler then uses Watson Text to Speech API to convert the dialogue into speech and tweaks the vocal features of the generated speech according to the OCEAN parameters of the agent. At the end of the agent speech, Scenario Handler checks whether the end state of the current scenario is reached, in which case it stops execution. Otherwise, it updates the state of the current scenario and starts a new turn.

Animation Modifier (see Figure 3) is responsible for adjusting the bone rotations and facial shape keys of the agent at each frame, by modifying the base animation according to agent personality. It maps the OCEAN values into three groups of animation modification parameters: Laban Shape Quality (LSQ), Laban Effort (LE), and Facial Expression. These parameters are calculated once at the beginning and used throughout the execution. LSQ parameters determine positions and weights of Inverse Kinematics (IK) targets; LE parameters further refine the animation with additional rotations; Facial Expression parameters regulate the interpolation of facial shape keys. We extract visemes from speech using Lipsync [Oculus 2019] and blend them into facial shape keys to animate the mouth. Agent emotions are shown as facial expressions and decay over time. We have developed the system in Unity [2019] and used Fuse [Adobe 2019] for creating 3D human models with a skeletal rig and facial shape keys.

To evaluate the system, we implemented three scenarios: Introduction, Fastfood, and Passport. Each scenario includes an appropriate scene setup, one main agent, and a dialogue state
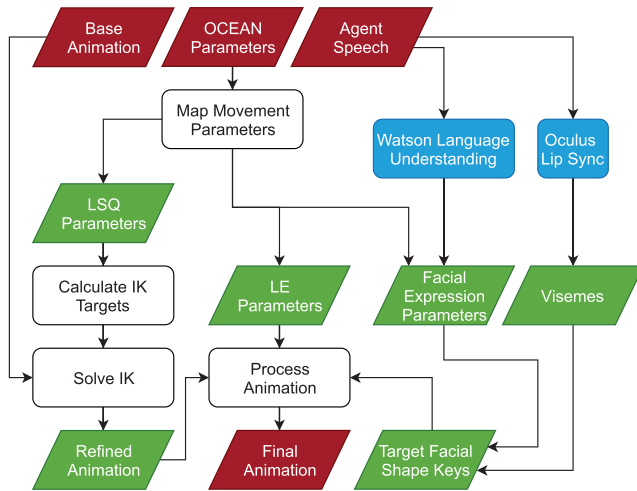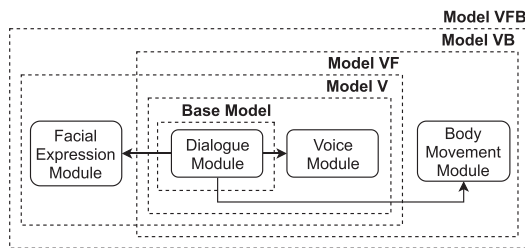
Fig. 3. The Animation Modifier flowchart.

Fig. 4. Personality modification modules and models used in the experiments.

machine. The scenarios aim to direct the user input into a specific direction, because each agent is designed for scenario-specific dialogues. The user is either given a task at the beginning of a scenario or guided with state-specific descriptions inside the scene. The scenarios are described in Appendix A. We performed a detailed perception study for the Passport Scenario. The Animation Modifier runs in real-time, and the bottleneck of the Scenario Handler is the external APIs that we use for speech transcription and generation. To minimize the latency caused by any external API, we used predefined user and agent sentences in the perception study.

4 PERSONALITY-DRIVEN MODIFICATION MODULES

We group the operations that modify a given input based on agent personality into four modules: dialogue, voice, body movement, and facial expression modification, as shown in Figure 4. Each module uses internal mappings to customize the animation for expressing the desired personality.

4.1 Dialogue Module

Generally, personality cues are inherent in dialogue content; an average person can make personality judgments based on dialogue text [Mairesse et al. 2007]. To our knowledge, except for domain-specific models [Mairesse and Walker 2010], the state-of-the-art natural language systems are not capable of generating general dialogue for OCEAN personality types [Dušek et al. 2020]. It is

Table 1. An Example for Personality Specific Dialogue Lines

Type	Text (Original: *"Ok, I will buy my return ticket."*)
$O(+)$	"Of course, I'm going to buy it for sure. I should have done it before; this will be a lesson for me."
$O(-)$	"Oh, ok. I will buy it."
$C(+)$	"Yes, I am going to buy my return ticket as soon as possible."
$C(-)$	"I will buy it... I will, as soon as possible."
$E(+)$	"Yes, I will buy the return ticket immediately. Thank you, officer."
$E(-)$	"Ok, I will."
$A(+)$	"Thank you very much for reminding, I'm going to buy it as soon as possible."
$A(-)$	"Well, I have to, you know. I will buy it."
$N(+)$	"I. . . I am going to buy it. I will buy it as soon as possible."
$N(-)$	"Sure, I will buy it as soon as possible."

a common approach to use handcrafted dialogue with embodied agents to express personality [Brixey and Novick 2019]. We leave the incorporation of an automated, personality-driven dialogue generation for future work and use handcraft utterances for each personality type, following the natural language features in Mairesse and Walker [2007].

Table 1 shows example variations for one line of dialogue from the Passport Scenario.

4.2 Voice Module

According to Polzehl [2015], the vocal features of speech can be adjusted to reinforce personality. Watson Text-to-Speech API [IBM 2015] supports multiple transformable voices. We use *"en_US_Michael"* for male and *"en_US_Allison"* for female voices. We use the following parameters of IBM Watson API for transforming the voice:

Pitch: The frequency of the voice. Speech sounds more excited when the pitch is high, and calmer when it is low.
Pitch range: The pitch variation during the speech. A higher range sounds more melodic and dynamic, and a lower range is monotonous.
Rate: The speed of the voice. A high rate is perceived as hurried and excited, and a low rate sounds calm.
Breathiness: The amount of escaping air during sound production. High breathiness is perceived as calm.
Glottal tension: The hardness of the voice. High glottal tension is perceived dynamic and tense, and low glottal tension sounds calmer and softer.

We follow the personality-to-vocal features mapping introduced by Polzehl [2015]. Polzehl uses words such as "high," "medium," and "low" in the mapping. We interpret these correlations as a weighted average of OCEAN factors as shown in Table 2. We map the feature values into different ranges to keep the transformed voice natural.

Table 2. Numerical Mapping between Vocal Features and OCEAN Personality, Following Polzehl [2015]

Feature	O	C	E	A	N	Range
Pitch	1	0.5	0.5	−0.5	−0.5	[−80, 80]
PitchRange	1	—	1	—	—	[−100, 100]
Rate	1	1	−1	1	—	[−70, 70]
Breathiness	1	—	—	1	—	[−50, 50]
GlottalTension	1	−1	—	1	—	[−100, 100]

Table 3. Additive Facial Expression Values from the Facial Expression-Personality Experiment

Factor	Hap.	Sad.	Ang.	Sur.	Dis.
O (+)	+0.13	0.00	−0.25	0.00	0.00
C (+)	+0.13	0.00	−0.13	−0.35	−0.13
E (+)	+0.25	−0.25	0.00	0.00	0.00
A (+)	+0.25	0.00	−0.50	−0.25	−0.35
N (+)	−0.13	+0.35	+0.50	+0.35	+0.35

Corresponding values are added to the agent's facial expression per turn of dialogue. Negative factors have the opposite sign.

4.3 Facial Expression Module

Recent studies have shown associations between facial features and judgment of personality on virtual characters [Ferstl and McDonnell 2018; Wang et al. 2013]. However, personality judgments of real and virtual faces tend to differ [Ferstl and McDonnell 2018], and the associations observed for real faces are not always consistent (e.g. correlating facial symmetry with high extraversion) [Fink et al. 2005], requiring further research. Evaluating our judgments and understanding possible biases are valuable, and employing these findings in character design can enhance our interaction with virtual characters. However, it also poses the risk of promoting stereotypes. So instead, we look for associations between personality and dynamic features of appearance, one of which is the usage of facial expressions.

We animate the facial expression of an agent, as a facade of its emotions, by blending facial shape keys calculated as a weighted sum of five different emotions: anger, sadness, happiness, disgust, surprise. We use 50 shape keys of the model generated from Adobe Fuse [Adobe 2019]. Please refer to Appendix B for the shape key weights of each emotion. We store the current and target values of shape keys in arrays and linearly interpolate the current value of each shape key toward its target at each frame.

Each emotion takes a value between 0 and 1 and decays over time. The emotion update is based on two inputs. The first one is the agent's current utterance, as dialogue is a natural indicator of emotional content for a conversational agent. The second one is the agent personality, which indicates a tendency to experience certain emotions more than others. We performed a user study to establish a link between personality perception and the facial expression of emotions. Please refer to Appendix C.2 for the details of this study.

Watson Natural Language Understanding API [IBM 2015] estimates emotions of anger, disgust, sadness, and happiness from text, returning an emotion value between 0 and 1 for each sentence. For example, "Such a lovely day." produces $joy = 0.78$, $anger = 0.01$, $disgust = 0.01$, $sadness = 0.03$, and $fear = 0.09$; while "I don't want to talk with you anymore" produces $joy = 0.04$, $anger = 0.11$, $disgust = 0.09$, $sadness = 0.59$, and $fear = 0.20$. For each line of agent dialogue, we update the agent's emotion value using the Watson API output for that line. Then, we add personality-based expression parameters given in Table 3 to align the expression of emotions with agent personality. For instance, for an extravert agent uttering "Such a lovely day," the facial expression parameters for happy, sad, angry, surprised, and disgusted will be updated by (0.78 + 0.25), (0.03−0.25), 0.01, 0.09, and 0.01, respectively. These values are then clamped between 0 and 1.

The resulting expression is a blend of multiple emotions [Martin et al. 2006]. Although some complex facial expressions can be interpreted differently when there are multiple blending factors, we do not focus on analyzing these combinations, because they rarely occur in the current system.

Emotion decay occurs while the agent is listening, which helps produce a natural listening animation [Maatman et al. 2005]. We calculate the emotion decay rate using neuroticism, similar to Kasap et al. [2009]. Additionally, we scale the facial expression by an expressiveness factor, which is determined by the extraversion parameter. This makes it possible to have subtle expressions on introvert agents and exaggerated ones on extraverts. Finally, different from the other shape keys, we update the blink speed based on neuroticism [Hoppe et al. 2018].

4.4 Body Movement Module

4.4.1 Movement Modifications by Laban Shape Qualities. Laban Shape Qualities describe the way the body changes form during movement in three orthogonal axes: *longitudinal* (Rising/Sinking), *frontal* (Spreading/Enclosing), and *sagittal* (Advancing/Retreating). We modify a given motion by orienting the end effectors (hands) toward a certain direction based on Shape parameters. To this end, we place six anchor points equidistantly around the center of the body, positioning them up, down, left, right, front, and back—the body center is the seventh anchor. These anchor points, along with Shape parameters, are used to update the hand IK targets at each frame.

We calculate Laban Shape Quality values ($LSQ_i \in [−1, 1]$, where $i \in \{Rising, Spreading, Advancing\}$) based on OCEAN parameters. The sign of LSQ_i determines the direction of movement. For example, $LSQ_{Rising} > 0$ indicates rising motion whereas $LSQ_{Rising} < 0$ indicates sinking motion. Each LSQ_i determines the attraction toward a specific anchor. For example, we use the top anchor for $LSQ_{Rising} > 0$, and the bottom anchor for $LSQ_{Rising} < 0$. Sagittal axis is computed similarly, using the top and bottom anchors for both hands. In the frontal axis, however, we need to use separate anchor points for each hand. We use the center anchor (Enclosing) and the left anchor (Spreading) for the left hand (Figure 5), and the center anchor (Enclosing) and the right anchor (Spreading) for the right hand.

We compute the attraction toward a specific direction of the corresponding axis by multiplying the sign of LSQ_i by an attraction factor. At each time step, we use linear interpolation to find IK target positions for each LSQ_i and take their average as the final IK
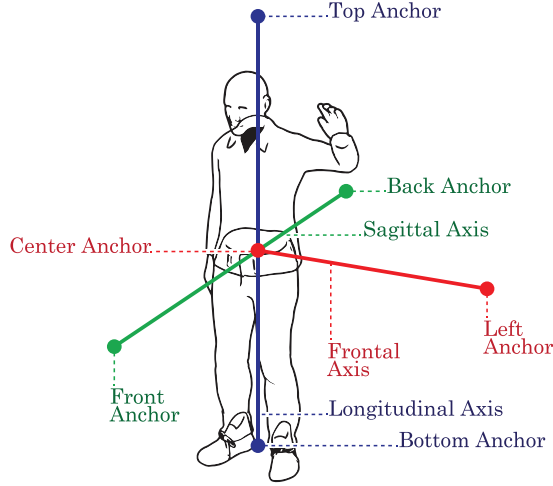
Fig. 5. The anchor points for the left hand.

Fig. 6. The trajectory of the left hand for a waving motion. The base animation is shown in red. The modified animations express Spreading Shape with a constant anchor weight (blue) and with variable anchor weights (green). Because waving has a wide range of arm movement, using constant anchor weights causes dramatic changes in the base animation, modifying its original curve and undermining its naturalness.

Fig. 7. The distances in attraction weight calculations, assuming h is the left hand, a is the left anchor, and the red arc is the path of the left hand during the motion. We project the distances on a line for better visualization.

target position for each hand h. We then feed the calculated positions into the IK Solver to generate a refined pose.

Depending on the nature of the animation, attraction factors can be constant or time-variant. As a preprocessing step, we calculate a constant attraction factor (AF) for each anchor, normalizing the difference between the farthest distance of the hand from the anchor and the closest distance of the hand from the anchor during the course of the hand's base animation. For certain motions, such as an idle animation where hands move slightly around their resting position, using this time-invariant attraction factor is sufficient. Because hand displacement is minimal in such cases, keeping the attraction to all the anchors at minimum avoids unnatural poses. However, larger-scale hand motions require the attraction factor to be updated dynamically. In those cases, we use time-variant attraction weights (W) to preserve the essence of the base animation. Multiplying them by AF gives us time-variant attraction factors AF_t. For instance, consider a pointing motion starting with the hand at a resting position and ending with the hand in a forward-pointing state. To update this motion in the sagittal direction, we need to move the hand position forward during the animation. Doing this at the starting state would impair the intent of pointing motion. Instead, the hand should be pushed further away from the body during the pointing step. In other words, we want to use the highest attraction weight toward the front anchor when the hand is closest to the anchor point and the lowest one when it is farthest from the anchor point. Figure 6 shows waving, which is a relatively large-scale hand motion, with constant and varying attraction weights. Constant anchor weights cause the hand trajectory to change dramatically, so the resulting motion no longer appears natural. In contrast, with variable anchor weights, the resulting animation is still similar to the base animation with Spreading Shape.

Although our current Laban Shape Quality adjustment algorithm is designed for atomic gestures, it is relatively straightforward to extend it to modify motion clips that are composed of multiple gestures, as long as we know the start and end frames of each gesture. We can then compute the hand movement range for each gesture, and update the attraction factors accordingly. Motion segmentation can be done manually, or better yet, automatically, considering the semantics of the movement. However, we leave this as a future work, because our scenario handling mechanism employs separate animation clips with atomic actions.

The calculation of attraction factors and weights is as follows:

$$NF(a, h, s) = \frac{1}{(AbsoluteMax(a, h, s) - AbsoluteMin(a, h, s))},$$

$$AF(a, h, m, s) = (max(d_{ahm}) - min(d_{ahm})) \cdot NF(a, h, s),$$

$$W(a, h, t, m, s) = \left(\frac{max(d_{ahm}) - d_{ahm}(t)}{max(d_{ahm}) - min(d_{ahm})} \right),$$

$$AF_t(a, h, t, m, s) = W(a, h, t, m, s) \cdot AF(a, h, m, s),$$

where $NF(a, h, s)$ is the normalizing factor for the skeleton s per anchor a and hand h. $AbsoluteMax(a, h, s)$ and $AbsoluteMin(a, h, s)$ functions find the absolute maximum and minimum a-to-h distances by projecting a line of arm length from the shoulder point of h away from anchor a, and toward anchor a, respectively (see Figure 7). $AF(a, h, m, s)$ is the attraction factor for hand h toward

Table 4. The Space Effort Rotation Axes and Signs

Bone	Rotation Axis	Space −	Space +
Left Shoulder	Longitudinal	+	−
Left Shoulder	Sagittal	−	+
Left Upper Arm	Sagittal	−	+
Left Foot	Longitudinal	+	−
Left Hand	Longitudinal	−	+
Left Fingers	Frontal	+	−

Bones on the right side have the opposite sign of the corresponding bone on the left side.

Table 5. The Weight Effort Rotation Axes and Signs

Bone	Rotation Axis	Weight −	Weight +
Spine	Frontal	−	+
Neck	Frontal	−	+
Left Shoulder	Sagittal	+	−
Left Upper Leg	Frontal	+	−
Left Lower Leg	Frontal	−	+

Bones on the right side have the opposite sign of the corresponding bone on the left side.

anchor a for motion m using skeleton s. It is a time-invariant factor that summarizes the movement range of motion m. d_{ahm} is the array of a-to-h distances for motion m, calculated per time step. $W(a, h, t, m, s)$ is the time-variant attraction weight of hand h toward anchor a at time t for motion m using skeleton s, which is multiplied by $AF(a, h, m, s)$ to get the time-variant attraction factor $AF_t(a, h, t, m, s)$.

4.4.2 Movement Modifications by Laban Effort Parameters. We map LE parameters into individual bone rotations that are blended into the current pose. We use the LE-OCEAN mapping introduced by Durupinar et al. [2017]. Because LMA concepts are reproducible using different motion parameters [Durupinar et al. 2017], we introduce a different implementation of LE modifications. We blend various bone rotations to add the impact of the Weight and Space components to the posture. We change animation speed dynamically to add the impact of the Time component, and we blend small rotations for each bone based on Perlin noise to add the impact of the Flow component.

The Space component reflects the inner attitude toward attention to the environment. Indirect Space (−) has an affinity with Spreading Shape, and Direct Space (+) has an affinity with Enclosing Shape. Blended rotation axes and signs for this component are shown in Table 4. We determine the rotation limits manually in a way that looks natural. The blending factor is determined by the magnitude of the Space parameter.

The Weight component reflects the gravitational force on the body. Blended rotation axes and signs for this component are shown in Table 5. Strong Weight (+) has a Sinking Shape, whereas Light Weight (−) has a Rising Shape.

We change the speed of the animation in a non-uniform manner to implement the Time Effort. We preprocess the base animation to determine a rank per time step, based on the average displacement of hands at that time step. During runtime, the animation speed is modified based on the rank of the corresponding time step and the

Time Effort value. The usage of variable speeds helps changing the movement speed in a more natural way.

As an example, consider a pointing gesture that starts and ends with the same idle pose. The hand transitions into the pointing position and stays still for a short time before returning to the idle pose. A direct speed-up would shorten the pointing time, but undermine realism. The preferred result is to make the transitional motion quick, without drastically changing the pointing duration. We preprocess the animation with a constant sample rate to calculate the average displacement of hands per step, following Algorithm 1. We rank each step according to the average hand displacement and map this to a speed factor. The speed factor is the highest for the time step with the largest hand displacement and the lowest for the one with the lowest hand displacement. The speed factor boundaries are adjustable.

ALGORITHM 1: Preprocessing for the animation speed update.

forall *Base Animation A* **do**
    DisplacementArr = new Float[A.FrameCount-1];
    for *t = 1 to A.FrameCount* **do**
        $d_L = GetHandDisplacementBwFrames(t-1, t, A.LeftHand)$;
        $d_R = GetHandDisplacementBwFrames(t-1, t, A.RightHand)$;
        $DisplacementArr[t] = (d_L + d_R) / 2$;
    end
    A.RankArr = new Integer[A.FrameCount-1];
    for *t = 1 to A.FrameCount* **do**
        A.RankArr[t] = GetRank(DisplacementArr[t]);
    end
end

We set the animation speed, which is an internal Unity parameter, during runtime using Algorithm 2. We determine the speed range [MinSpeed: MaxSpeed] by the Time Effort parameter, which is in the $[−1, 1]$ range. For $Time = −1$, the speed range is $[0.5 : 1]$; and for $Time = 1$, it is $[1 : 2]$. The resulting animation appears faster or slower without breaking relative stops.

ALGORITHM 2: Setting the animation speed at each frame.

A = GetCurrentBaseAnimation();
t = A.GetCurrentFrame();
rank = A.RankArr[t];
A.speed = Map rank from [1, A.FrameCount] to [MinSpeed, MaxSpeed];

We add rotational fluctuations per bone using Perlin noise based on the Flow Effort parameter to have a similar effect to the flourishes in Durupinar et al. [2017]. Free Flow has maximum fluctuations, indicating uncontrollability; whereas Bound Flow has no additional rotational fluctuations. Figure 8 depicts the influence of Laban Effort parameters on the agent's body.

For OCEAN-LE mapping, we use the mapping introduced by Durupinar et al. [2017]. As for the OCEAN-LSQ mapping, no direct mathematical link has been defined in the literature to the best of our knowledge. Therefore, we follow an experimental approach to quantize the OCEAN-LSQ link considering the descriptions of Shape Qualities and hand movement in body language [Key 1975]. For example, an extravert tends to expose his/her hands more than an introvert, this is why we associate positive extraversion with Opening LSQ. LSQ parameters are calculated as a weighted sum

(a) Direct Space/Indirect Space   (b) Light Weight/Strong Weight

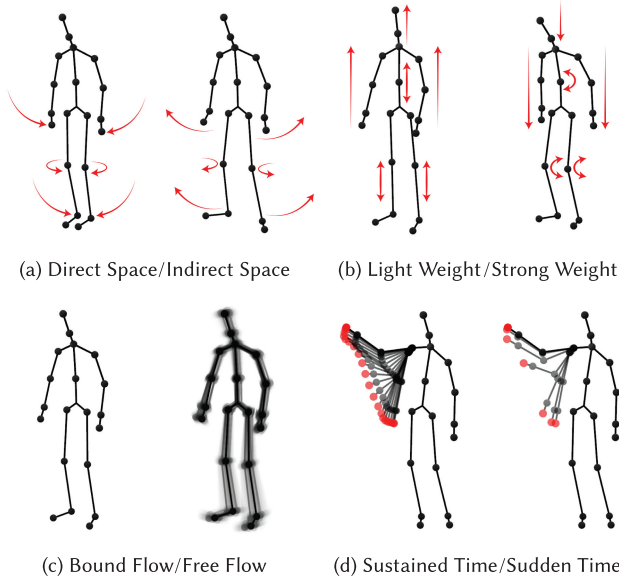(c) Bound Flow/Free Flow   (d) Sustained Time/Sudden Time

Fig. 8. The influence of each Laban Effort modification. (a) Enclosing vs. spreading motion for Direct vs. Indirect Space. (b) Rising vs. sinking motion for Light vs. Strong Weight. (c) Constrained vs. fluctuating movements for Bound vs. Free Flow. Figures show the superimposition of multiple frames of a still motion. (d) Slow vs. fast movement for Sustained vs. Sudden Time. Figures show the superimposition of multiple frames of a pointing motion. The red dots illustrate the hand position captured at each time step. Because Sustained Time has slow movement, it spans more frames than Sudden Time.

Table 6. OCEAN Weights for LSQ

LSQ/Personality	O	C	E	A	N
Rising	0.25	0.25	0.25	0.25	0.00
Spreading	0.30	0.00	0.70	0.00	0.00
Advancing	0.00	0.00	0.50	0.00	0.50

of OCEAN factors, using the weights in Table 6. All the movement and personality parameters take values in the range $[-1, 1]$. Through perceptual studies, we verified that the Shape adjustments contribute to the perception of personality and improve the distinction of personality factors compared to solely using Effort-based modifications. Please refer to Appendix C.3 for the experimental evaluation.

5 EXPERIMENTS

5.1 Experimental Setup

We performed perception studies on Mechanical Turk [Amazon 2018] to evaluate the amount of distinction between the two extremes of each personality trait, using incremental combinations of personality expression modules. The system can express each OCEAN factor on a scale and combine the influence of different factors. However, we focus on singular factor changes with the extreme values to keep the number of experiments at a manageable size.

We performed two preliminary experiments to tune various system parameters before the main experiment. The first preliminary experiment is a user survey to select the most neutral looking character to prevent personality perception bias due to the appearance of the 3D model. Please refer to Appendix C.1 for details. The second one is another user study to test the effect of facial expressions on the perception of personality and construct a link between the two. To our knowledge, emotions have not been used as a direct element to influence the perceived personality of a virtual character. We provide the details in Appendix C.2 and use the results of the experiment in our personality to facial expression mapping. In addition to the preliminary experiments, we ran two user studies to (1) validate the Laban modifications for body motion and (2) evaluate the naturalness of the agent in terms of its movement and speech. The details are provided in Appendices C.3 and C.4.

We performed eight main experiments, wherein we tested the same 3D agent model with different active communication channels. Each main experiment assessed ten samples of the model endowed with one OCEAN extreme at a time while the remaining factors were set to neutral. All ten samples of each experiment were rated by 50 unique participants. We did not allow the participants to take part in multiple experiments to avoid familiarity bias. Thus, a total of 400 Turkers participated in the eight main experiments, where non-overlapping sets of 50 Turkers rated all the ten samples of each experiment. We conducted the main experiments in two stages. The difference between the two stages is the presence/absence of dialogue and voice. The first stage involved dialogue, where vocal communication was active in all but one case, and the text was available in all. The second stage disabled both the text (dialogue) and voice modules to keep the attention on body and face movements. The second stage experiment samples were shorter, focusing only on the key actions. We presented the samples in random order. We ran the Passport Scenario, which covers the interaction between a passport officer (the user representative) and a visitor (the agent) always with the same scenario flow, except for the agent answers, which were crafted to reflect the desired personality. We showed the video of each conversation, expecting answers to personality statements about the agent on a 7-point Likert scale. We used the Ten-Item Personality Inventory [Gosling et al. 2003] for the statements, which were in the following structure: "This character looks <u>Extraverted</u> and <u>Enthusiastic</u>," with the underlined traits updated depending on the personality type. Each sample and the accompanying statements were displayed on the same page. For each sample, we randomized agent names and occupations to make sure that they would be perceived as different individuals. Participants were free to view each sample as many times as they needed. We did not have control over the setup where the participants performed the studies, as they were run on Mechanical Turk. Participants were not able to return to a sample after submitting their response. Providing demographics was not mandatory. Among the 64% of all participants who shared this information, 73.79% are male, 27.21% are female, and the average age is 31.46 ± 7.64. The majority of the participants are from India (43.14%), followed by USA (25.80%), Spain (14.11%), and England (5.64%).

Table 7. Modules in Each Model

Model	Dialogue	Voice	Face	Body
Base Model	+	−	−	−
Model V	+	+	−	−
Model VF	+	+	+	−
Model VB	+	+	−	+
Model VFB	+	+	+	+
Model F	−	−	+	−
Model B	−	−	−	+
Model FB	−	−	+	+

"+" denotes active and "−" denotes inactive.

5.2 Models

Each main experiment involves a model with different active modalities. Table 7 shows the active and inactive modules in each model, and Figure 9 shows screenshots from each model in the first stage.

5.3 Results and Analysis

For each experiment, we performed Welch's t-tests to compare the means of the participant scores for positive and negative samples, where the null hypothesis assumes no significant difference. Because we made multiple comparisons, we adjusted the p-values to correct Type 1 error using False Discovery Rate control [Benjamini and Hochberg 1995].

Tables 8 and 9 show the mean differences between the scores of positive and negative samples followed by the adjusted p-value of each mean difference per OCEAN factor. The columns represent the pairs of simulations with two personality poles and the models. The magnitude of the difference indicates the clarity of the distinction between the positive and negative samples, demonstrating the success of the modifications made for that personality trait. A small p-value shows that the difference is not coincidental.

5.3.1 Perception of Individual OCEAN Factors. For openness, the perception of the Models VB and VFB yield significant mean differences between the negative and positive extremes. Combining facial expressions with movement cues slightly increases the distinction factor– the mean difference is higher for VFB. However, we observe that dialogue, voice, and body movement play the main role here, as Model VF does not yield a significant difference by itself. In the absence of dialogue and voice, none of the models achieve a significant perceptual difference between the two extremes of openness, alluding to the intellectual associations of this trait.

For conscientiousness, Models VF, VB, and VFB have significant effects on helping distinguish the two extremes. The combination of VF and VB seems to have an averaging effect on VFB. Individually, Models V and FB have similar performance, suggesting that their combination has an additive influence. Conscientiousness is best distinguished by Model VB. Similar to openness, the absence of dialogue and voice result in poor performance.

Extraversion is significantly distinguished from introversion using the models V, VB, VFB, F, and FB, where VB and VFB perform the best. Among all the OCEAN factors, extraversion is the best-distinguished one from its opposite, with a mean difference of 2.96

(a) Positive Extraversion Sample (Base Model and Model V)

(b) Positive Extraversion Sample (Model VF)

(c) Negative Extraversion Sample (Model VF)

(d) Positive Extraversion Sample (Model VB)

(e) Negative Extraversion Sample (Model VB)
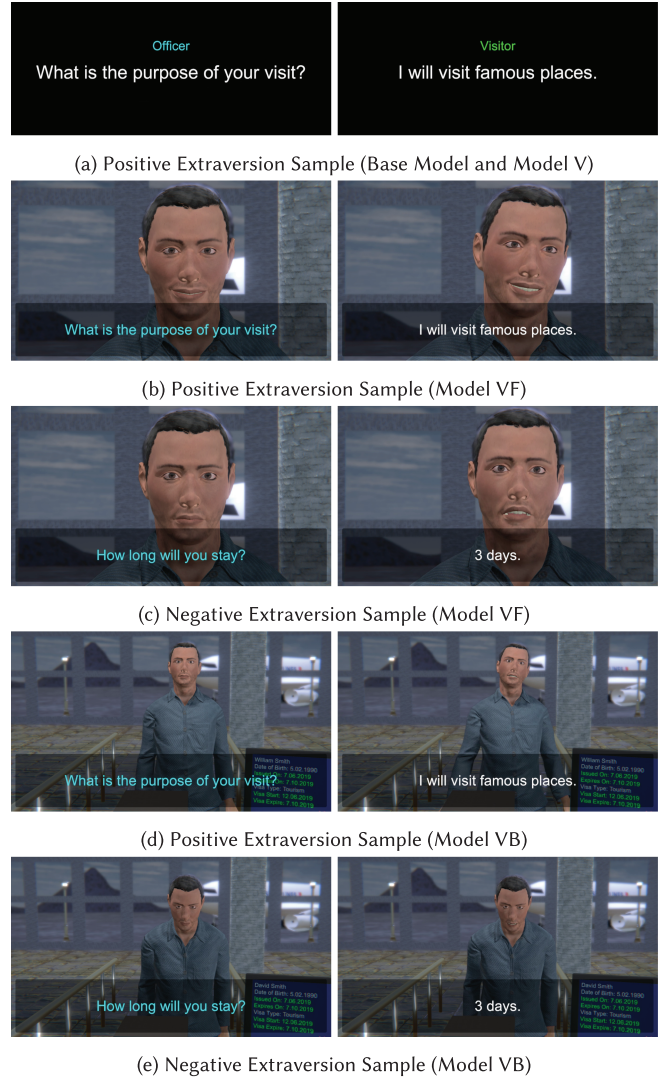
Fig. 9. Screenshots showing each model, excluding Model VFB, which is similar to Model VB, except that the facial expressions are as in Model VF, rather than being neutral. Among the second stage experiments, Model F is similar to Model VF, Model B is similar to Model VB, and Model FB is similar to Model VFB, excluding dialogue and voice.

using Model VFB. It is also the one that gains the most benefit from combining facial expressions with body movement modifications. Interestingly, for extraversion, although Model VB outperforms Model VF, the exclusion of dialogue and voice makes Model F to perform better than Model B. This is possibly due to the change of focus between the two stages of experiments. The first stage is longer, shifting viewers' attention toward dialogue and body, decreasing the influence of facial expressions.

The two poles of agreeableness can be distinguished significantly using Models V, VF, VB, VFB, F, and FB. The best performing models are VB, VFB, and FB. Using only body movement or dialogue is not adequate to achieve significant differences in means. The best distinguishable factor using Model V is agreeableness,

Table 8. OCEAN Score Difference and Significance of This Difference per Model and Target Factor

	Sample O+, O−					Sample C+, C−					Sample E+, E−					Sample A+, A−					Sample N+, N−				
	Base	V	VF	VB	VFB	Base	V	VF	VB	VFB	Base	V	VF	VB	VFB	Base	V	VF	VB	VFB	Base	V	VF	VB	VFB
Δ_O	0.47	0.55	0.85	1.16	1.53	0.24	0.12	−0.06	0.46	0.56	−0.14	0.71	0.85	1.20	1.32	0.27	0.20	0.91	0.67	0.86	−0.29	−0.31	−0.96	−.68	−1.19
ρ_O	0.002	0.036	0.001	0.000	0.000	0.045	0.002	0.000	0.000	0.000	0.204	0.000	0.000	0.000	0.000	0.011	0.016	0.001	0.000	0.000	0.939	0.001	0.000	0.000	.000
Δ_C	0.37	0.87	1.17	1.48	1.92	0.38	0.82	1.18	1.74	1.50	0.03	0.31	0.61	0.61	0.73	0.22	0.58	0.86	0.98	1.54	−0.56	−1.07	−1.88	−2.62	−2.74
ρ_C	0.193	0.522	0.775	0.002	0.003	0.057	0.002	0.000	0.000	0.000	0.562	0.315	0.478	0.053	0.002	0.873	0.167	0.657	0.030	0.085	0.016	0.000	0.035	.000	.000
Δ_E	0.14	1.39	1.17	2.00	2.63	0.10	0.19	0.15	0.44	0.82	0.16	0.94	1.56	2.12	2.96	0.12	0.40	0.66	0.16	0.66	−0.45	−0.59	−1.26	−1.15	−1.64
ρ_E	0.296	0.004	0.002	0.000	0.000	0.826	0.086	0.004	0.000	0.000	0.342	0.003	0.000	0.000	0.000	0.241	0.004	0.001	0.000	0.000	0.999	0.242	0.001	.002	.000
Δ_A	0.47	0.68	1.14	1.03	1.71	−.03	0.27	0.09	0.44	0.37	0.14	0.66	0.76	0.86	1.01	0.62	1.11	1.81	2.32	2.65	−0.26	−0.25	−0.71	−.45	−.74
ρ_A	0.131	0.353	0.000	0.002	0.000	0.203	0.015	0.009	0.000	0.000	0.394	0.054	0.007	0.561	0.009	0.009	0.001	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000
Δ_N	−0.01	−0.88	−1.24	−1.11	−1.55	−0.43	−0.96	−0.58	−1.35	−1.49	0.00	−0.22	−0.84	−0.78	−1.22	−0.48	−1.14	−1.61	−2.01	−2.63	0.58	1.05	2.04	2.14	2.76
ρ_N	0.093	0.138	0.001	0.000	0.000	0.025	0.002	0.000	0.000	0.000	0.033	0.018	0.000	0.000	0.000	0.040	0.253	0.005	0.028	0.001	0.013	0.000	0.000	.000	.000

Δ_x represents the mean score difference (in terms of the 7 point likert-scale) between $y+$ and $y−$ samples for each sample $y \in \{O, C, E, A, N\}$, per dimension $x \in \{O, C, E, A, N\}$, rounded to 2 decimals. ρ_x represents the adjusted p-value for the significance of the difference between $y+$ and $y−$ samples in terms of dimension x, rounded to 3 decimals. Columns represent the two extreme samples of the dimension of interest. Sub-columns represent each model that is used in the experiments. Gray cells indicate $\rho_x < 0.05$ with $1 \le \Delta_x < 2$, blue cells indicate $\rho_x \le 0.001$ with $1 \le \Delta_x < 2$, and green cells indicate $\rho_x \le 0.001$ with $2 \le \Delta_x$.

Table 9. OCEAN Score Difference and Significance of This Difference per Silent Model and Target Factor

	Sample O+, O−			Sample C+, C−			Sample E+, E−			Sample A+, A−			Sample N+, N−		
	B	F	FB	B	F	FB	B	F	FB	B	F	FB	B	F	FB
Δ_O	0.20	0.47	0.34	0.04	0.24	−0.07	0.12	1.14	0.96	0.45	1.34	0.81	−0.14	−0.78	−0.60
ρ_O	0.440	0.069	0.182	0.870	0.320	0.794	0.643	0.000	0.000	0.084	0.000	0.001	0.595	0.005	0.033
Δ_C	0.27	1.16	0.71	0.29	0.63	0.80	0.15	0.94	0.36	0.65	1.08	1.24	−0.09	−1.29	−1.08
ρ_C	0.368	0.000	0.013	0.315	0.033	0.004	0.626	0.000	0.214	0.016	0.001	0.000	0.745	0.000	0.000
Δ_E	0.42	0.01	0.50	−0.26	−0.02	−0.46	0.82	1.31	1.95	0.56	1.05	0.98	−0.17	−0.86	−0.88
ρ_E	0.130	0.969	0.067	0.386	0.946	0.128	0.006	0.000	0.000	0.051	0.001	0.001	0.585	0.001	0.003
Δ_A	−0.15	1.35	1.45	−0.23	0.49	0.43	0.40	1.11	1.00	0.42	1.91	2.03	−0.22	−0.79	−1.11
ρ_A	0.625	0.000	0.000	0.437	0.118	0.128	0.207	0.000	0.001	0.118	0.000	0.000	0.434	0.010	0.000
Δ_N	−0.24	−1.83	−1.67	−0.12	−0.78	−1.16	−0.52	−1.33	−0.91	−0.65	−1.85	−2.11	0.31	1.99	2.30
ρ_N	0.465	0.000	0.000	0.691	0.009	0.001	0.089	0.000	0.005	0.032	0.000	0.000	0.318	0.000	0.000

Please refer to Table 8 caption for details.

namely, vocal features contribute to the perception of agreeableness more than the other personality traits. Despite the apparent influence of voice, when dialogue and voice are excluded, Model FB still achieves a large mean difference. This suggests that facial expressions and vocal features may be substitutable for agreeableness.

Most of the models represent neuroticism well. Between neuroticism and stability, Models V, VF, VB, and VFB yield significant mean differences, with Model V slightly less than the others. Among these, Model VFB performs the best. When voice and dialogue are enabled, body movement is as effective as facial expressions in expressing neuroticism. However, excluding the dialogue and voice has a dramatic negative influence on the effect of body movement.

Some traits require longer exposure times so that the viewer can assess their influence on perceived personality. Features such as the dependability of conscientiousness and the adventurous nature of openness can be hardly illustrated in short sequences. Behavioral planning similar to Shvo et al. [2019] can be used to express such traits through the scenario flow. Additionally, the second stage experiments indicate that viewer attention is controlled by the existence of dialogue and voice, as well as the exposure duration. In general, longer sequences with dialogue shift the attention toward the agent's body, whereas short sequences without dialogue keep the focus on facial expressions, reducing the performance of body movement modifications. Providing a reference point, such as comparing two agents, as in our movement modifications evaluation (see Appendix C.3), increases the influence of body movement.

5.3.2 Comparison of Models. We compare the results of the models by performing one-way ANalysis Of VAriance (ANOVA) with Tukey Honestly Significant Difference (HSD) to reveal significant differences for each OCEAN factor [Tukey 1949]. Tables 10 and 11 show the difference-in-differences (DID) in means of pairs of models, which are based on positive and negative samples for each OCEAN factor. Each value reports the improvement provided by the model on the left over the model on the right. For example, Model VFB is 2.80 times more successful than Base Model for extraversion. Note that we evaluate the dialogue-based and silent models separately as they involve different experimental setups. The information in these tables is also presented as graphs in Appendix D.

In the dialogue-based experiments, combining facial expressions with movement modifications works best for extraversion; probably because each module adds non-overlapping features

Table 10. The DID in Means of the Pairs of Models for Each OCEAN Factor

Model	O	C	E	A	N
V - Base	0.08	0.44	0.78	0.49	0.47
VF - Base	0.38	0.80	**1.40**	1.19	**1.46**
VB - Base	0.69	**1.36**	**1.96**	**1.70**	**1.56**
VFB - Base	**1.06**	**1.12**	**2.80**	**2.03**	**2.18**
VF - V	0.30	0.36	0.62	0.70	0.99
VB - V	0.61	0.92	**1.18**	1.21	1.09
VFB - V	**0.98**	0.68	**2.02**	**1.54**	**1.71**
VB - VF	0.31	0.56	0.56	0.51	0.10
VFB - VF	0.68	0.32	**1.40**	0.84	0.72
VFB - VB	0.37	−0.24	0.84	0.33	0.62

Bold values show significant differences ($p < 0.05$).

Table 11. The DID in Means of the Pairs of Silent Models for Each OCEAN Factor

Model	O	C	E	A	N
F-B	**0.27**	0.34	0.49	**1.49**	**1.68**
FB-B	0.14	0.51	**1.13**	**1.61**	**1.99**
FB-F	−0.13	0.17	0.64	0.12	0.31

Bold values show significant differences ($p < 0.05$).

related to extraversion. For instance, facial expressions project positive emotions and sociability, whereas body movement conveys the energetic nature of extraversion. We observe a similar effect on neuroticism. Again, we can surmise that each modality contributes to a different aspect of neuroticism expression.

Without dialogue and voice, the face becomes more important than the body in expressing personality. This is demonstrated in openness, agreeableness, and neuroticism. A combination of facial expressions and body motion has clear advantages over only body motion for extroversion, agreeableness, and neuroticism. However, adding body motion to facial expressions yields no significant improvements over only adjusting facial expressions.

5.3.3 Correlation of OCEAN Factors. We observe correlations between the perception of different OCEAN factors in different models. For instance, the modifications that focus on openness increase the perception of extraversion. The adjustments to express conscientiousness tend to be negatively correlated with the perception of neuroticism. Such an outcome is expected, because slower movements convey conscientious behavior, which can be perceived as calm and stable—an indicator of negative neuroticism. For the first stage of experiments, where dialogue and voice are enabled, we perform principal component analysis (PCA) for Models VF, VB, and VFB to analyze the correlations. We ignore the Base Model and Model V, because they do not have enough separation. We apply PCA to the models individually, where the population consists of OCEAN scores of each sample rated by each participant. This results in matrices of 500 rows (10 samples × 50 participants) and 5 columns (each OCEAN factor) per model.

We use Kaiser-Meyer-Olkin (KMO) test [Dziuban and Shirkey 1974] as a measure to show the suitability of the data for PCA. KMO greater than 0.7 is generally accepted as an adequate threshold for factor analysis. Each principal component in Table 12 acts

Table 12. PCA Results for Models VF, VB, and VFB

Model	VF			VB			VFB		
Comp.	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
O	0.65	—	0.68	0.70	0.49	0.37	0.77	0.35	—
C	0.74	—	−0.37	0.83	—	−0.38	0.81	—	−0.42
E	0.62	0.67	—	0.68	0.57	—	0.71	0.55	0.31
A	0.72	-0.53	—	0.72	−0.45	0.45	0.76	−0.39	0.45
N	−0.83	—	—	−0.84	0.35	—	−0.88	—	—
CVar	50.71	66.80	81.40	57.14	75.52	86.56	62.43	75.92	86.63
KMO	0.752			0.747			0.810		

KMO shows Kaiser-Meyer-Olkin test result, we suppress coefficients less than 0.3, and display components that capture at least 80% of the cumulative variance (CVar).

as a new linearly uncorrelated dimension that captures some portion of the variance. We specify the coefficients for each OCEAN factor per principal component, suppressing values less than 0.3. The results show that we can use three components to capture at least 80% of the cumulative variance for models VF, VB, and VFB.

Each positive OCEAN dimension except neuroticism has at least one trait that has positive connotations in a general sense. For instance, imagination (O), dependability (C), sociability (E), politeness (A), and calmness (N-) are among the descriptive traits of the five factors. We expect that people's perception of personality is influenced by a dimension that encapsulates the overall positiveness, similar to the Big One in the literature [Musek 2007]. PC1 in all models can be interpreted as this positiveness dimension, which is exhibited the most in positive conscientiousness and negative neuroticism.

We interpret PC2 as the dominant extraversion dimension. In general, extraversion is the best-distinguished OCEAN factor. We believe that this is due to the expanding and energetic behavior that it encompasses. Other than extraversion, PC2 is correlated with different factors in each model. In Model VF, it is not surprising that PC2 is related to positive extraversion (happiness expression) and negative agreeableness (anger and disgust expressions) where strong and energetic facial expressions emerge. In Model VB, extraversion is coupled well with openness, similar to Smith and Neff [2017], but it also related to negative agreeableness and positive neuroticism. The open posture of positive openness, extending limbs of high extraversion, quick and angry movements of low agreeableness, and rushed motion of high neuroticism all contribute to this energetic behavior component. We believe that in Model VFB, the inclusion of facial expressions helps distinguish whether the motion speed signals energetic or anxious behavior.

PC3 yields a complementary output to PC2, with the common correlation factor being negative conscientiousness across models. In Model VF, it can be interpreted as a creativity dimension where highly creative—a feature of positive openness−individuals generally express negative conscientiousness traits such as being less organized [Feist 2019].

In Model VB, PC3 consists of negative conscientiousness, positive openness, and agreeableness. This result may be due to the less bound and faster movements of negative conscientiousness. In Model VFB, the additive influence of facial expressions for extraversion and agreeableness may be responsible for the additional variance. Again, more relaxed use of expressions such as smiling

Table 13. PCA Results for Models F and FB, with Similar Setup as Table 12

Model	F			FB		
Comp.	PC1	PC2	PC3	PC1	PC2	PC3
O	0.80	—	0.45	0.69	0.48	—
C	0.81	—	−0.31	0.75	−0.35	0.45
E	0.50	0.83	—	0.47	0.78	—
A	0.82	—	—	0.77	—	−0.52
N	−0.85	—	—	−0.84	0.31	—
Cvar	58.95	77.05	86.68	51.08	73.31	84.68
KMO	0.767			0.703		

may be indicators of a more carefree attitude, as seen in negative conscientiousness. This carefree attitude would go well with the cheerful nature of positive extraversion and forgiving mindset of positive agreeableness, which would explain the correlation.

PCA on the second stage experiments reveals similar results (Table 13). We exclude Model B, since it yields $KMO < 0.7$. PC1 in both models expresses a similar Big One-like structure, where extraversion is not dominant. PC2 again represents the dominant extraversion dimension, coupled with negative conscientiousness in Model FB. Interestingly, PC2 in Model F has no other contributing factors. PC3 in Model F is similar to PC3 in Model VF, with openness being less influential due to the exclusion of dialogue and voice. In model FB, the order of negative conscientiousness and negative agreeableness dimensions are switched, compared to Model VFB. This can be explained by dialogue improving the distinction of agreeableness. Overall, PCA results are very similar to the first stage experiments, only with a reduced representation of openness and agreeableness.

6 DISCUSSION

The experiments indicate that adding new modalities improves personality perception, and the combination of all the modalities performs the best. Although adding a new modality does not always yield a statistically significant increase in the perceptual accuracy, the overall effect is additive. There is only one exception where the improvement slightly decreases when a new modality is added, and that is for conscientiousness when body movement is added to facial animation and voice. However, the amount of the decrease is not statistically significant. When the communication channels are all active, all the five personality factors show statistically significant improvements over the baseline model, where personality is expressed only through dialogue. Previous research also indicates the effectiveness of multi-modal communication. For instance, Neff et al. show that combining language variation and gesture parameters improves the perception of extraversion [Neff et al. 2010] and neuroticism [Neff et al. 2011] in conversational agents. In another domain, for personality detection from videos of real people, combining speech signals with hand and body movement descriptors performs better than individual channels [Nguyen et al. 2013].

We verify that the users perceived the models as natural in a survey described in Appendix C.4. The survey results suggest that the users judged the speech based on the naturalness of body motion and facial expressions. The users reported a slight decrease in naturalness when facial expressions and body movement modifications were combined; however, they generally agree that agent movement, facial expression, and voice modifications are natural.

Although we can conclude that multi-modal communication improves personality perception in general, there are some variations across the contributions of different combinations depending on the personality factor. For example, the combination of dialogue and voice modules performs slightly better than using body and face modules together for openness and conscientiousness. In contrast, the body-face combination outperforms the dialogue-voice combination for the remaining factors. Thus, in systems where dialogue is not prominent, such as crowd simulations, we expect extroversion, agreeableness, and neuroticism to be conveyed more effectively. This is indeed backed by previous crowd simulation work [Durupinar et al. 2011], where the recognition rates of these three factors are higher than the rates of openness and conscientiousness.

Among the five factors, extraversion is distinguished the best, followed by neuroticism and agreeableness. These three factors are more expressive as they are related to social attitudes. In fact, the markers for extraversion, neuroticism, and agreeableness are found to be more consistent across different species [Gosling and John 1999]. Openness and conscientiousness, however, are related to the intellectual and long-term aspects of personality [Goldberg 1990]; and they are not distinguished as much. We believe that capturing their cues would require exposure to the agent in different contexts and for longer durations. Their especially poor performance in the absence of dialogue implies the need for contextual cues to represent them.

The results show that the modifications that focus on one OCEAN factor also influence the perception of other OCEAN dimensions, although they are considered orthogonal. Understanding the reason behind these correlations requires further experiments, but one possible explanation is the limited number of movement modifications to represent each factor. For instance, movement speed has different interpretations for neuroticism and conscientiousness [Durupinar et al. 2017], which we found to be inversely correlated. Sudden movements are perceived as erratic, and thus the agent looks neurotic. At the same time, they imply carelessness, so observers associate them with low conscientiousness. However, it is interesting to see the same relationship between neuroticism and conscientiousness when only the voice module is active. Several studies have identified two meta-traits above the five OCEAN dimensions through subsequent factor analysis [Cieciuch and Strus 2017; DeYoung et al. 2002]. The two meta-traits group extraversion with openness, and conscientiousness with agreeableness and neuroticism. These are consistent with the correlations that we found. For example, Table 8 shows similar results for openness and extraversion, albeit with a higher recognition rate for extraversion. However, there is a clear distinction between extraversion and agreeableness, which belong to different meta-traits.

7 CONCLUSION AND FUTURE RESEARCH

We combine various methods of personality expression in conversational virtual agents to create a multi-modal framework and

evaluate the individual and combined influence of different modalities on personality perception. Body animation, facial expressions, voice transformation, and dialogue content are used in cooperation to convey the underlying psychological states of the agent to the user. The user studies suggest that each additional module contributes to the perception of personality, with the best performance achieved when all the modules are active. Some limitations and future research directions are as follows:

Real-time conversation with the agent: We performed the perception studies using recordings of the agent to prevent potential delays due to the external APIs. Using the current system, there is a 2–6 s delay after the user's speech until the agent starts speaking. We hope to reduce this delay to an acceptable range by using local speech generation and understanding techniques. In the future, we plan to test the system with users directly interacting with the agent through conversation. We expect direct interaction to increase user engagement in the scenario and elicit more empathy toward the agent. It would also be interesting to observe how users adjust their conversation styles based on agent personality. For instance, users might talk more politely to a highly agreeable agent, tend to keep their conversation short with an introverted agent, and be impatient with an unconscientious agent.

Personality-based automated style transfer for text: Currently, we use handcrafted dialogue; however, injecting personality into neutral sentences can provide richer dialogue. This would require large data sets with sentences that have a similar meaning and different personality scores for training. Having multiple units that focus on singular text operations is a possible solution. For example, one unit could be responsible for adding hesitation into text by a given degree, another for adding surprise, and so on.

Data-driven approach: Using motion capture data with personality scores, a mapping between animation modification parameters and personality can be defined. Similar to Ran et al. [2015] who introduce an automated model to predict LMA qualities of recorded motion, deep learning can be used to modify an animation based on personality. Facial expression data can also be used to accompany movement modifications.

A more sophisticated model for emotions: Emotions are currently reflected as facial expressions that are customized based on the agent's dialogue sentences and personality. A more sophisticated model involves agent emotions to be influenced by the different aspects of communication with the user, such as acting defensively against an angry user. An emotion model similar to Kasap et al. [2009] can be utilized for this purpose. A complex emotion model also entails the influence of personality. For example, a highly agreeable agent would be expected to be more caring when the user sounds sad, as opposed to a disagreeable agent that would not be interested in cheering up the user. In addition, we currently limit the communication of emotions to facial expressions. Although the link between personality traits and body motion has been established in the literature, we have little knowledge about how this link would be affected when emotions are superimposed on personality. This is a future research area that we aim to explore.

Interaction with multiple agents: Currently, the system focuses on conversation with a single agent. It could be interesting to integrate it into a simulation where multiple agents interact based on personality. The goals and needs for each agent could be determined by a model, and human players could train agent behavior.

REFERENCES

Kfir Aberman, Yijia Weng, Dani Lischinski, Daniel Cohen-Or, and Baoquan Chen. 2020. Unpaired motion style transfer from video to animation. *ACM Trans. Graph.* 39, 4 (2020), 64.

Barbara Adrian. 2008. *Actor Training the Laban Way: An Integrated Approach to Voice, Speech, and Movement.* Allworth, New York, NY.

Jan Allbeck and Norman Badler. 2002. Toward representing agent behaviors modified by personality and emotion. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, Vol. 2. ACM, New York, NY, 15–19.

Gene Ball and Jack Breese. 2001. *Emotion and Personality in a Conversational Agent.* MIT Press, Cambridge, MA, 189–219.

Ligia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. 2013. Cicero—Towards a multimodal virtual audience platform for public speaking training. In *Intelligent Virtual Agents.* Springer, Berlin, 116–128.

Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc., Series B Stat. Methodol.* 57, 1 (1995), 289–300.

Aniket Bera, Tanmay Randhavane, and Dinesh Manocha. 2017. Aggressive, tense or shy? Identifying personality traits from crowd videos. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17).* IJCAI, San Mateo, CA, 112–118.

Ran Bernstein, Tal Shafir, Rachelle Tsachor, Karen Studd, and Assaf Schuster. 2015. Laban movement analysis using kinect. *Int. J. Comp. Inf. Eng.* 9 (2015), 1394–1398.

Andrew Best, Sahil Narang, and Dinesh Manocha. 2020. SPA: Verbal Interactions between Agents and Avatars in Shared Virtual Environments using Propositional Planning. Retrieved from https://arxiv:cs.MA/2002.03246.

Jacqueline Brixey and David Novick. 2019. Building rapport with extraverted and introverted agents. In *Advanced Social Interaction with Agents.* Springer, Cham, Switzerland, 3–13.

Stéphanie Buisine, Améziane Aoussat, and Jean-Claude Martin. 2007. Embodied creative agents: A preliminary social-cognitive framework. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA'07).* Springer-Verlag, Berlin, 304–316.

Shanna L. Burke, Tammy Bresnahan, Tan Li, Katrina Epnere, Albert Rizzo, Mary Partin, Robert M. Ahlness, and Matthew Trimmer. 2017. Using virtual interactive training agents (ViTA) with adults with autism and other developmental disabilities. *J. Autism Dev. Disord.* 48, 3 (Nov. 2017), 905–912.

Sarah Jane Burton, Ali-Akbar Samadani, Rob Gorbet, and Dana Kulić. 2016. Laban movement analysis and affective movement generation for robots and other near-living creatures. In *Dance Notations and Robot Motion.* Springer, Cham, Switzerland, 25–48.

Justine Cassell, Catherine Pelachaud, Norman I. Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. 1994. Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of SIGGRAPH'94.* ACM, New York, NY, 413–420.

Ginevra Castellano, Maurizio Mancini, Christopher Peters, and Peter W. McOwan. 2011. Expressive copying behavior for social agents: A perceptual analysis. *IEEE Trans. Sys., Man Cybern. Part A: Syst. Hum.* 42, 3 (2011), 776–783.

Constantinos Charalambous, Zerrin Yumak, and A. Frank van der Stappen. 2019. Audio-driven emotional speech animation for interactive virtual characters. *Comp. Anim. Virt. Worlds* 30, 3–4, Article e1892 (2019), 11 pages.

Jan Cieciuch and Włodzimierz Strus. 2017. Two-factor model of personality. In *Encyclopedia of Personality and Individual Differences*, Virgil Zeigler-Hill and Todd K. Shackelford (Eds.). Springer, Cham, 1–17.

Aline W. de Borst and Beatrice de Gelder. 2015. Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Front. Psychol.* 6 (2015), 576.

Colin G. DeYoung, Jordan B. Peterson, and Daniel M. Higgins. 2002. Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Pers. Individ. Differ.* 33, 4 (2002), 533–552.

Funda Durupınar, Uğur Güdükbay, Aytek Aman, and Norman I. Badler. 2016. Psychological parameters for crowd simulation: From audiences to mobs. *IEEE Trans. Vis. Comp. Graph.* 22, 9 (2016), 2145–2159.

Funda Durupinar, Mubbasir Kapadia, Susan Deutsch, Michael Neff, and Norman I. Badler. 2017. PERFORM: Perceptual approach for adding OCEAN personality to human motion using laban movement analysis. *ACM Trans. Graph.* 36, 1, Article 6 (2017), 16 pages.

F. Durupinar, N. Pelechano, J. Allbeck, U. Güdükbay, and N. I. Badler. 2011. How the Ocean personality model affects the perception of crowds. *IEEE Comput. Graph. Appl.* 31, 3 (2011), 22–31.

Ondřej Dušek, Jekaterina Novikova, and Verena Rieser. 2020. Evaluating the state-of-the-art of end-to-end natural language generation: The E2E NLG challenge. *Comput. Speech Lang.* 59 (2020), 123–156.

Charles D. Dziuban and Edwin C. Shirkey. 1974. When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychol. Bull.* 81, 6 (1974), 358.

Paul Ekman. 2019. The 5 Communication Channels. Retrieved from https://www.ekmaninternational.com/wiki/the-5-communication-channels/.

Hugo Jair Escalante, Heysem Kaya, Albert Ali Salah, Sergio Escalera, Yagmur Gucluturk, Umut Guclu, Xavier Baró, Isabelle Guyon, Julio Jacques Junior, Meysam Madadi et al. 2018. Explaining first impressions: Modeling, recognizing, and explaining apparent personality from videos. Retrieved from https://arXiv:1802.00745 abs/1802.00745.

Hans J. Eysenck and Michael W. Eysenck. 1985. *Personality and Individual Differences: A Natural Science Approach.* Plenum Press, New York, NY.

Gregory J. Feist. 2019. Creativity and the Big Two model of personality: Plasticity and stability. *Curr. Opin. Behav. Sci.* 27 (2019), 31–35.

Ylva Ferstl and Rachel McDonnell. 2018. A perceptual study on the manipulation of facial features for trait portrayal in virtual agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA'18).* ACM, New York, NY, 281–288.

Bernhard Fink, Nick Neave, John T. Manning, and Karl Grammer. 2005. Facial symmetry and the "big-five" personality factors. *Pers. Individ. Differ.* 39, 3 (2005), 523–529.

Patrick Gebhard. 2005. ALMA—A layered model of affect. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'05).* ACM, New York, NY, 29–36.

Leilani H. Gilpin, Danielle M. Olson, and Tarfah Alrashed. 2018. Perception of speaker personality traits using speech signals. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI EA'18).* ACM, New York, NY, Article LBW514, 6 pages.

Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI EA'11).* ACM, New York, NY, 253–262.

Lewis R. Goldberg. 1990. An alternative "Description of Personality": The big-five factor structure. *J. Pers. Soc. Psychol.* 59 (1990), 1216–1229.

Samuel D. Gosling and Oliver P. John. 1999. Personality dimensions in nonhuman animals: A cross-species review. *Curr. Dir. Psychol. Sci.* 8, 3 (1999), 69–75.

Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann Jr. 2003. A very brief measure of the big-five personality domains. *J. Res. Personal.* 37, 6 (2003), 504–528.

Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents.* Springer, Berlin, 125–138.

Björn Hartmann, Maurizio Mancini, and Catherine Pelachaud. 2002. Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis. In *Proceedings of the Conference on Computer Animation (CA'02).* IEEE, Piscataway, NJ, 111–119.

Sabrina Hoppe, Tobias Loetscher, Stephanie A. Morey, and Andreas Bulling. 2018. Eye movements during everyday behavior predict personality traits. *Front. Hum. Neurosci.* 12, Article 105 (2018), 8 pages.

Zerrin Kasap, Maher Ben Moussa, Parag Chaudhuri, and Nadia Magnenat-Thalmann. 2009. Making them remember-emotional virtual characters with memory. *IEEE Comput. Graph. Appl.* 29, 2 (2009), 20–29.

Mary Ritchie Key. 1975. *Paralanguage and Kinesics (Nonverbal Communication).* Scarecrow Pres, Lanham, MD.

Caroline Larboulette and Sylvie Gibet. 2015. A review of computable expressive descriptors of human motion. In *Proceedings of the 2nd International Workshop on Movement and Computing (MOCO'15).* ACM, New York, NY, 21–28.

Jina Lee and Stacy Marsella. 2006. Nonverbal behavior generator for embodied conversational agents. In *Proceedings of the International Workshop on Intelligent Virtual Agents (IVA'06).* Springer, Heidelberg, 243–255.

Tino Lourens, Roos Van Berkel, and Emilia Barakova. 2010. Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis. *Rob. Auton. Syst.* 58, 12 (2010), 1256–1265.

Gale M. Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Comput. Hum. Behav.* 37 (2014), 94–100.

R. M. Maatman, Jonathan Gratch, and Stacy Marsella. 2005. Natural behavior of a listening agent. In *Intelligent Virtual Agents*, T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist (Eds.). Springer, Berlin, 25–36.

François Mairesse and Marilyn Walker. 2007. PERSONAGE: Personality generation for dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computer Linguistics.* Association of Computing Linguistics, Stroudsburg PA, 496–503.

François Mairesse and Marilyn A. Walker. 2010. Towards personality-based user adaptation: Psychologically informed stylistic language generation. *User Model User-adapt Interact.* 20, 3 (2010), 227–278.

François Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Art. Intell. Res.* 30 (2007), 457–500.

Vera Maletic. 2011. *Body-Space-Expression: The Development of Rudolf Laban's Movement and Dance Concepts.* Vol. 75. Walter de Gruyter, Berlin.

Jean-Claude Martin, Radoslaw Niewiadomski, Laurence Devillers, Stéphanie Buisine, and Catherine Pelachaud. 2006. Multimodal complex emotions: Gesture expressivity and blended facial expressions. *Int J. Hum. Robot.* 3 (09 2006), 269–291.

Megumi Masuda and Shohei Kato. 2010. Motion rendering system for emotion expression of human form robots based on laban movement analysis. In *Proceedings of the 19th International Symposium on Robot Human Interaction and Communication (RO-MAN'10).* IEEE, Piscataway, NJ, 324–329.

Adobe. 2019. Fuse CC. Retrieved from https://www.adobe.com/tr/products/fuse.html.

Amazon. 2018. Mechanical Turk. Retrieved from https://www.mturk.com/.

IBM. 2015. IBM Watson API. Retrieved from https://www.ibm.com/watson.

Oculus. 2019. Oculus Lipsync. Retrieved from https://developer.oculus.com/downloads/package/oculus-lipsync-unity/.

Unity. 2019. Unity. Retrieved from https://unity.com/.

Robert R. McCrae and Paul T. Costa. 2005. *Personality in Adulthood: A Five-factor Theory Perspective.* Guilford Press, New York, NY.

Robert R. McCrae and Oliver P. John. 1992. An introduction to the five-factor model and its applications. *J. Pers.* 60, 2 (1992), 175–215.

Rachel McDonnell, Martin Breidt, and Heinrich H. Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.* 31, 4, Article Article 91 (July 2012), 11 pages.

Janek Musek. 2007. A general factor of personality: Evidence for the Big One in the five-factor model. *J. Res. Pers.* 41, 6 (2007), 1213–1233.

Laura P. Naumann, Simine Vazire, Peter J. Rentfrow, and Samuel D. Gosling. 2009. Personality judgments based on physical appearance. *Pers. Soc. Psyc. Bull.* 35, 12 (2009), 1661–1671.

Michael Neff, Nicholas Toothman, Robeson Bowmani, Jean E. Fox Tree, and Marilyn Walker. 2011. Don't scratch! Self-adaptors reflect emotional stability. In *Proceedings of the Conference on Intelligent Virtual Agents (IVA'11), LNAI.* Springer-Verlag, Berlin, 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 *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA'10).* Springer, Berlin, 222–235.

Laurent Son Nguyen, Alvaro Marcos-Ramiro, Martha Marrón Romera, and Daniel Gatica-Perez. 2013. Multimodal analysis of body communication cues in employment interviews. In *Proceedings of the 15th ACM International Conference on Multimodal Interaction (ICMI'13).* ACM, New York, NY, 437–444.

Truong-Huy D. Nguyen, Elin Carstensdottir, Nhi Ngo, Magy Seif El-Nasr, Matt Gray, Derek Isaacowitz, and David Desteno. 2015. Modeling warmth and competence in virtual characters. In *Intelligent Virtual Agents*, Willem-Paul Brinkman, Joost Broekens, and Dirk Heylen (Eds.). Springer, Cham, 167–180.

Catherine Pelachaud. 2005. Multimodal expressive embodied conversational agents. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MM'05).* ACM, New York, NY, 683–689.

Tim Polzehl. 2015. *Personality in Speech: Assessment and Automatic Classification.* Springer, Cham, Switzerland.

Bernstein Ran, Shafir Tal, Tsachor Rachelle, Studd Karen, and Schuster Assaf. 2015. Multitask learning for laban movement analysis. In *Proceedings of the 2nd International Workshop on Movement and Computation (MOCO'15).* ACM, New York, NY, 37–44.

Tanmay Randhavane, Aniket Bera, Kyra Kapsaskis, Kurt Gray, and Dinesh Manocha. 2019a. FVA: Modeling perceived friendliness of virtual agents using movement characteristics. *IEEE Trans. Vis. Comp. Graph.* 25, 11 (2019), 3135–3145.

Tanmay Randhavane, Aniket Bera, Kyra Kapsaskis, Rahul Sheth, Kurt Gray, and Dinesh Manocha. 2019b. EVA: Generating emotional behavior of virtual agents using expressive features of gait and gaze. In *Proceedings of the ACM Symposium on Applied Perception (SAP'19).* ACM, New York, NY, Article Article 6, 10 pages.

Tanmay Randhavane, Aniket Bera, Emily Kubin, Kurt Gray, and Dinesh Manocha. 2019c. Modeling Data-Driven Dominance Traits for Virtual Characters using Gait Analysis. Retrieved from https://arxiv:cs.GR/1901.02037.

Maayan Shvo, Jakob Buhmann, and Mubbasir Kapadia. 2019. An interdependent model of personality, motivation, emotion, and mood for intelligent virtual agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA'19).* ACM, New York, NY, 65–72.

Harrison Jesse Smith and Michael Neff. 2017. Understanding the impact of animated gesture performance on personality perceptions. *ACM Trans. Graph.* 36, 4, Article 49 (2017), 15 pages.

Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psyc.* 29, 1 (2010), 24–54.

Angela Tinwell, Mark Grimshaw, Debbie Abdel Nabi, and Andrew Williams. 2011. Facial expression of emotion and perception of the uncanny valley in virtual characters. *Comput. Hum. Behav.* 27, 2 (March 2011), 741–749.

Alexander Todorov, Chris P. Said, Andrew D. Engell, and Nikolaas N. Oosterhof. 2008. Understanding evaluation of faces on social dimensions. *Trends Cogn. Sci.* 12, 12 (2008), 455–460.

John Tukey. 1949. Comparing individual means in the analysis of variance. *Biometrics* 5, 2 (1949), 99–114.

Marco Vala, Gabriel Blanco, and Ana Paiva. 2011. Providing gender to embodied conversational agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA'11)*. Springer-Verlag, Berlin, 148–154.

Bart van Straalen, Dirk Heylen, Mariët Theune, and Anton Nijholt. 2009. *Enhancing Embodied Conversational Agents with Social and Emotional Capabilities*. Springer, Berlin, 95–106.

Y. Wang, J. Geigel, and A. Herbert. 2013. Reading personality: Avatar vs. human faces. In *Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction*. 479–484. DOI : 10.1109/ACII.2013.85

Shihong Xia, Congyi Wang, Jinxiang Chai, and Jessica Hodgins. 2015. Realtime style transfer for unlabeled heterogeneous human motion. *ACM Trans. Graph.* 34, 4, Article Article 119 (July 2015), 10 pages.

Zerrin Yumak and Nadia Magnenat-Thalmann. 2016. Multimodal and multi-party social interactions. In *Context Aware Human-Robot and Human-Agent Interaction*, Nadia Magnenat-Thalmann, Junsong Yuan, Daniel Thalmann, and Bum-Jae You (Eds.). Springer, Cham, 275–298.

Mehmet Ersin Yümer and Niloy J. Mitra. 2016. Spectral style transfer for human motion between independent actions. *ACM Trans. Graph.* 35, 4, Article 137 (2016), 8 pages.

Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To stylize or not to stylize? The effect of shape and material stylization on the perception of computer-generated faces. *ACM Trans. Graph.* 34, 6, Article Article 184 (Oct. 2015), 12 pages.

Liwei Zhao and Norman I. Badler. 2001. *Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures*. Technical Report IRCS-01-11. University of Pennsylvania.

Michelle X. Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting virtual agents. *ACM Trans. Interac. Intell. Sys.* 9, 2–3 (Mar. 2019), 1–36.