Investigating How Speech And Animation Realism Influence The Perceived Personality Of Virtual Characters And Agents

Sean Thomas* Technological University Dublin
Ylva Ferstl† Trinity College Dublin
Rachel McDonnell‡ Trinity College Dublin
Cathy Ennis§ Technological University Dublin

Figure 1: Examples from our stimuli featuring combined voice, motion and appearance modalities through each of our characters. From left to right illustrates Females 1-3 followed by Males 1-3.

Abstract

The portrayed personality of virtual characters and agents is understood to influence how we perceive and engage with digital applications. Understanding how the features of speech and animation drive portrayed personality allows us to intentionally design characters to be more personalized and engaging. In this study, we use performance capture data of unscripted conversations from a variety of actors to explore the perceptual outcomes associated with the modalities of speech and motion. Specifically, we contrast full performance-driven characters to those portrayed by generated gestures and synthesized speech, analysing how the features of each influence portrayed personality according to the Big Five personality traits. We find that processing speech and motion can have mixed effects on such traits, with our results highlighting motion as the dominant modality for portraying extraversion and speech as dominant for communicating agreeableness and emotional stability. Our results can support the Extended Reality (XR) community in development of virtual characters, social agents and 3D User Interface (3DUI) agents portraying a range of targeted personalities.

Index Terms: Embodied agents, virtual humans and (self-)avatars—Perception and cognition——

1 Introduction

Virtual characters and agents digitally represent humans in a variety of contexts, such as in computer games and motion pictures, as virtual tutors [44], streamers [59], medical practitioners [8], and 3D virtual assistants[1]. These characters are important for standalone and interconnected ("metaverse") 3D virtual worlds [16], which in addition to entertainment, are expected to accommodate future virtual currencies, businesses, jobs, laws and properties [6]. As these platforms flourish, it is important for Extended Reality (XR) innovators to understand how personality portrayal may be impacted by verbal and nonverbal features, as this could unintentionally dampen or abstract the perceived personality of individuals represented by avatars for embodied interactions in virtual environments - potentially misrepresenting them. Alternatively, users may be enabled to endow their avatars with custom characteristics - whether reflective of their own personality or not. Verbal [12, 35] and nonverbal [19, 48, 67] cues of virtual agents can have notable impacts on how we perceive them and understanding how modality fidelity impacts the perceived agent personality can support the development of targeted 3DUI agent personalities, enhancing agent affability in educational, medical or social XR applications.

Nonverbal communication as a means of expressing personality is highly effective; facial expressions, body language and qualities of a person's voice can be used as reliable indicators of personality [27]. Impressions of some traits, such as extraversion, can be formed with minimal short-term information [39], whereas other traits such as openness and conscientiousness relate to longer or repeated exposure [24]. Prior works demonstrate character personality expression through modalities such as positioning [10], facial expressions [57], gestures [47], voice [9], dialogue [57] and appearance [71]. Expectations of virtual assistant personalities appear to align with a blend of human-like and machine-like characteristics [51], resulting in a broad spectrum of requirements for personality expression. In this paper, we explore the impacts of verbal and nonverbal communication cues on the portrayed personality of virtual characters. We explore how the modalities of speech and motion impact personality portrayal for naturally conversing humans, and how personality is retained or altered when the fidelity of these modalities is reduced. Using datasets containing full-body motion capture and synchronous voice recordings of unscripted conversations, we compare motion capture and voice recordings to state-machine-like gesture animations with synthesized Text-to-Speech (TTS) audio, reflecting the typical animation and voice shortcomings of embodied conversational agents (ECAs). We find differing impacts of modalities on the Big Five personality traits [28], indicating that changes in modality fidelity can have both heightening and dampening effects on personality. We find that high motion fidelity is particularly important for communicating extraversion, whereas lower fidelity may increase perceived conscientiousness. We also find that agreeableness and emotional stability rely heavily on the speech modality. Our results offer insights that may assist in creating targeted personalities for interactive agents. To the best of our knowledge, this is the first work to explore how voice and motion degradations impact the portrayed

*e-mail: sean.a.thomas@mytudublin.ie
†e-mail: yferstl@tcd.ie
‡e-mail: ramcdonn@tcd.ie
§e-mail: cathy.ennis@tudublin.ie

[1] https://www.soulmachines.com/

personality of virtual characters and ECAs for unscripted conversations. Furthermore, our stimuli feature actors and avatars with a diverse range of accents and appearances - differing from many works including only an individual character of a single sex.

2 Related Work

Personality judgements are demonstrated to influence the trust, behaviour, and interactions that people share with virtual characters and agents [3, 49, 69]. These judgements may be influenced by characteristics such as appearance, body motion, facial motion, and voice [68]. Previous work compares the varying levels of realism between humanoid and robotic characters according to voice, motion and appearance [22], with research suggesting that increased modalities improve an agent's performance at portraying personality [57], indicating that adding more nonverbal features may directly improve the accuracy of personality portrayal. Nonverbal features such as body shape, attractiveness and posture contribute to our interpretation of personalities [62] in addition to verbal features such as vocalization patterns and prosody [62]. Prosody is particularly important, as inadequate prosody can impact naturalness [18], in turn influencing perceived personality. Isbister and Nass [26] observe that people prefer characters whose personalities compliment their own during interaction. However, other work notes that people did not show a greater rapport when interacting with agents most matching their level of extraversion [7]. These dissimilar findings suggest ambiguity in the understanding of how personality affects our perception of characters and agents. The content of speech and choice of wording is also a potential influence, as work suggests language utterances may alter the perceived degree of extraversion [42]. Aylett et al. observe that for most traits, personality attributions to synthetic voices are *"not an assessment of naturalness"* [4], suggesting that personality from speech is less related to authenticity and more related to the speaker's distinct characteristics. Pan et al. [49] found different behavioural responses to contrasting character personalities, even when verbal content is identical.

An important consideration is the first impression formed upon initial encounters with an agent. Impressions of agents attributed to facial features, proximity and gaze may form within the first 12.5 seconds of interaction [10]. Additionally, people may infer prompt judgements from brief exposure to faces [64], recognize emotion rapidly and effectively through voice [40] and consistently perceive personality through brief one-word utterances [43]. This poses the question of which verbal and nonverbal features most contribute to our initial impressions of personality.

Pennebaker and King [50] suggest that linguistic styles meaningfully contribute to personality, and that both positive and negative correlations exist between emotional words and unique personality traits, indicating that a person's choice of wording may contribute to personality judgements. Understanding the linguistic style of the text used to generate nonverbal behaviours may enable better implementations of personality-driven animations. The analysis and annotation of text [29, 30, 37] and the use of markup languages [15, 60] are widely explored for nonverbal behaviour generation. Physical appearance is another notable factor, with work showing impressions manifesting from both static (e.g. healthy, distinctive) and dynamic (e.g. smiling, energetic stance) appearance cues [46], aligning with other personality-related work observing perceptual implications for virtual characters with a non-healthy (*"ill"*) appearance [71].

Previous work has employed Laban Movement Analysis (LMA) for creating personality-driven character or agent behaviours. Chi et al. created a 3D character animation system, EMOTE, that generates synthetic gestures according to the *Effort* and *Shape* qualities of LMA [11]. Durupinar et al. [17] map low-level motion parameters to LMA parameters to the Big Five personality traits, enabling personality-driven expressive motion synthesis. They found that injecting personality can create more powerful expression, but at the cost of animations appearing cartoonish. Smith and Neff [56] investigated how modifying gesture motion affects perceived personality, finding that perceptions of personality could be modelled through two rather than impacting five dimensions, *"plasticity"* and *"stability"*. They found perceptions of extraversion and openness to be captured by plasticity, and agreeableness, conscientiousness and emotional stability by stability. Other work suggests that an interplay between personality and various factors (motivation, emotion and mood) may facilitate better interactions between humans and agents [55]. Agents participating in non-scripted conversations will require complex solutions to accommodate such interplaying factors in real-time. This is a challenging process, as agents must simultaneously interpret and react to conversational partners whilst responding plausibly via their output modalities.

Literature indicates that perceptual judgements are brief but accurate [43, 64], and that both verbal [45, 52] and nonverbal [5, 27] features make significant contributions to our impressions of others' personalities - particularly when such features are presented in combination. In this work, we contrast and examine the influence of motion and speech realism on personality. We explore how realism impacts perceptual outcomes for virtual characters and agents according to each Big Five personality trait; [28] (i) extraversion, (ii) agreeableness, (iii) conscientiousness, (iv) emotional stability and (v) openness to experience. Related work has explored the effects of modalities according to their individual and combined influence for scenario-specific interactions [57]. Similar our work is that of Koppensteiner et al. [32] who investigated personality based on full and reduced channels using videos of politicians' speeches, finding that vocal information dominated impression formation throughout all conditions - possibly related to politicians speaking in a more intentional, conscientious manner that is less natural and interpersonal. In our paper, we also assess full and reduced channels, but with an emphasis on modality fidelity in unscripted, natural conversations to understand how personality portrayal is altered by varying levels of motion and voice realism.

Table 1: Reference to the implemented performance capture dataset recordings. This includes the Talking With Hands [36] and Trinity Speech-Gesture 2 datasets [21].

Actor	Dataset	Session	Take(s)
Female 1	Talking With Hands	Session 32	7, 12
Female 2	Talking With Hands	Session 21	4
Female 3	Talking With Hands	Session 29	6, 7
Male 1	Talking With Hands	Session 23	11
Male 2	Trinity Speech Gesture 2	Session 1	7
Male 3	Talking With Hands	Session 24	18

3 Stimuli Creation

Our stimuli contain a set of informal conversational scenarios, in which different characters represented by distinct avatars speak to the viewer. Each character is created from the performance capture of a unique actor. Below, we describe how data was obtained and processed for our perceptual studies.

3.1 Data Selection

We obtained synchronized motion capture and audio data from the Talking With Hands [36] and Trinity Speech-Gesture 2 [21] datasets. We searched each dataset to identify segments that satisfied the following speech criteria: (i) the speaker is clear and intelligible, (ii) the chosen speech segment surpasses 30 seconds, (iii) the audio contains no significant microphone-leaking or interruptions from non-primary speakers and (iv) the dialogue context is unaffected

by masked-out vocabulary. Next, we assessed the quality of motion capture data for these speech segments and noted all viable recordings for our use case. We identified six speakers (3 female, 3 male) who best matched our criteria and cleaned the respective data. Data cleaning consisted of applying a 4Hz Butterworth filter to each motion capture file to reduce noise, normalizing the volume levels between each audio recording, and removing any unwanted background noise. We transcribed each audio recording to written text in preparation for our subsequent Text-to-Speech (TTS) conversion. To maintain consistency between recordings, we masked idle hand motion onto all of our stimuli. For replicability and to allow future work to select personality-specific samples based on our results, our chosen recordings from each dataset are referenced in Table 1.

3.2 Text-to-Speech

We passed our text transcriptions to Amazon Polly's standard engine[2] to generate TTS audio for our characters, making use of the standard engine in order to have access to a larger variety of synthetic voices. Rather than generating TTS for non-native speakers of English using the English output language, we adhered to Gluszek and Dovidio's suggestion that the inclusion of accents may *"help to develop a positive in-group identity"* and *"attenuate negative effects of perceived discrimination"* [23]. Using Amazon Polly's accented language tags from their Speech Synthesis Markup Language (SSML) documentation [3], we generated speech using non-English voices (chosen to best match each actor) to produce English output, resulting in accented English with pronunciation traits of the voices' native languages. The resulting audio files for all speakers were edited and aligned to match the key timings and co-speech gestures of the original performance.

3.3 Robotic Motion

We sought to mimic the animation style resulting from a state-machine-type animation method whereby segments of animation are selected and inserted for each part of an agent's performance. We created animations that retained the original co-speech gesture information of an actor's performance, but removed all auxiliary information such as body posture, head motion and between-gesture motion. Specifically, we retained the motion-captured arm and hand data of the stroke phase of each gesture [31] (motion signal automatically annotated using the classifier of Ferstl et al. [20]) and synthesized gesture transitions using software based on the DANCE animation environment [54] using spline interpolation. The remaining body (legs, torso, head) was animated with idle motion, resulting in a robotic motion style that retained nonverbal information from the co-speech gestures.

3.4 Character Selection

We created a set of six avatars (3 female, 3 male) using the Ready Player Me[4] platform to represent each performance capture actor. Character appearance has significant contributions to the overall judgement and perception of virtual characters and agents [14,61,70]. We ensured that clothing, eye color and eyebrow shape were homogenized between characters, as previous works show that clothing [38] and facial features [33,41,65] influence the perception of personality traits. In our appearance-only experiment (Section 4.2), we test the perceived personality across our chosen character appearances.

3.5 Animation and Rendering

Stimuli were created using Unity's High Definition Render Pipeline (HDRP), with a custom scene featuring a background from Poly Haven[5] and integrating subtle post-processing such as film grain with a depth of field filter (Figure 1). Identical shaders were applied to all characters for consistency. Videos were exported at 1920x1080 (100fps) using the Unity Recorder plugin, then compressed to 1280x720 (30fps) for an improved bitrate for playback in our experiment system. In line with similar work [22], we procedurally generated a single set of intervals for eye-blinks that we mapped uniformly to each character. An exception to this rule was to accommodate motion capture performances where the speaker rotates their head away from the in-engine camera, for which we encoded temporary remappings of eye gaze and used additionally encoded eye-blinks as transitions to give the appearance of our speakers naturally breaking eye-contact with the participant. Lip synchronization was implemented using the Oculus Lipsync for Unity plugin.

[2] https://aws.amazon.com/polly/
[3] https://docs.aws.amazon.com/polly/
[4] https://readyplayer.me/
[5] https://polyhaven.com/

Table 2: Notations describing experiment conditions.

	Voice (V)	**Motion (M)**
$V_N M_N$	*Natural* Voice (V_N)	*Natural* Motion (M_N)
$V_N M_R$	*Natural* Voice (V_N)	*Robotic* Motion (M_R)
$V_R M_N$	*Robotic* Voice (V_R)	*Natural* Motion (M_N)
$V_R M_R$	*Robotic* Voice (V_R)	*Robotic* Motion (M_R)

4 Experiment Design

We designed five perceptual experiments to study the effects and interactions of motion, voice, appearance and language (as text) on the portrayed personality of our characters. We explore each modality in isolation (Experiments I-IV), followed by a combination of all modalities (Experiment V). We denote our conditions in the format $V_X M_X$, where V refers to the voice and M to the motion component (Table 2). Experiments with only one modality are represented by V_X or M_X. Voice and Motion each have two condition expressions, *Natural* and *Robotic*. For Voice, the *Natural* condition (V_N) represents the audio from the original performance capture sessions, and *Robotic* (V_R) represents synthesized Text-to-Speech (TTS) audio generated from transcriptions of the performance capture sessions (Section 3.2). For Motion, the *Natural* condition (M_N) represents full-body performance capture with idle animations masked to the hands and fingers. *Robotic* (M_R) represents performance capture of the gesture stroke phase with synthesized gesture transitions and idle animation masked to the remainder of the body (Section 3.3). Our multimodal experiment conditions are listed in Table 2.

Participants were recruited via Prolific with English fluency required and paid according to Prolific's recommended hourly rate. Each participant engaged in only one experiment, and was instructed to view and rate a single practice clip prior to viewing and rating all stimuli. For experiments including audio or video, attention checks were embedded in the playback of stimuli and featured a multiple choice question related to the actor's appearance or the content of their speech. For appearance and text experiments, attention checks were embedded in the pre-experiment instructions and post-experiment questions to ensure that participants understood the task. Data of participants who failed an attention check was excluded. Experiments I-II contained 6 static stimuli (1 per actor) and lasted 10 minutes with no break, while Experiments III-V contained 12 video/audio stimuli (2 per actor, median duration of 51.5 seconds) and lasted 30 minutes with an optional midway 5-minute break. Participants were presented one stimulus at a time, each followed by a prompt to rate the presented character according to the Ten Item Personality Inventory (TIPI) [25]. The corresponding 7-point Likert scales ranged from "Disagree strongly" to "Agree strongly". The 10-item scores were collapsed to the Big Five's 5-item scores by averaging the primary ratings (e.g. *Extraverted, Enthusiastic*) with the reverse score of the contrary ratings (e.g. *Reserved, Quiet*). At the end of the experiment, participants were invited to provide

free-form feedback about their observations.

4.1 Experiment I - Text-only

We assessed how an actors' speech content may portray personality by utilizing text transcriptions of each actor's speech. Based on findings by Pennebaker et al. [50] for emotional language, we hypothesized that extraversion, agreeableness and emotional stability may be communicated through the text. (19 participants (10F, 9M), ages 19-51 years ($\mu = 28$, $\sigma = 7.96$)).

4.2 Experiment II - Appearance-only

Each of our actors was portrayed by a unique avatar. To assess any potential impact of our chosen character appearances, we conducted a controlled experiment in which participants rated the personality of characters based on a static image. Each character was presented in the same pose. Due to our efforts to maintain consistency for key visual characteristic across avatars (Section 3.4), we did not expect substantial differences in personality ratings solely from appearance. However, we did hypothesize that character gender would provoke variance for personality. (19 participants (9F, 10M), ages 18-68 years $\mu = 29$, $\sigma = 12.5$)).

4.3 Experiment III - Voice-only

We hypothesized that TTS would be more monotonous and would therefore dampen the perception of personality traits, but less so for traits displayed heavily through motion, such as extraversion (Section 4.5). This experiment used the *Natural* Voice (V_N) and *Robotic* TTS (V_R) conditions with no motion (i.e. only audio). (19 participants (11F, 8M), ages 19-35 years ($\mu = 26$, $\sigma = 4.5$)).

4.4 Experiment IV - Motion-only

We hypothesized that for motion-only presentation, the *Robotic* motion condition would have a particularly strong impact on the personality trait communicated heavily through this modality, namely extraversion (Section 4.5). This experiment compared the *Natural* Motion (M_N) and *Robotic* Motion (M_R) conditions. (19 participants (8F, 11M), ages 20-32 years ($\mu = 24$, $\sigma = 3.2$)).

4.5 Experiment V - Combined Modalities

Based on previous research, we hypothesized that audio would be more influential on perceived agreeableness and conscientiousness, whereas motion would strongly influence extraversion [56, 57]. For emotional stability and openness to experience, previous research suggest the interplay of both speech and motion would be important [56, 57]. Furthermore, we hypothesized that lower-realism motion and voice conditions would have a dampening effect on the perceived personality of each character. This experiment included all conditions from Table 2. (38 participants (19F, 18M, 1GQ), ages 18-47 years ($\mu = 24$, $\sigma = 5.8$)). Half of the participants (19 total) viewed all mocap-animated characters twice; once with natural voice and once with TTS. The other half (19 total) viewed the same but with robotic-animated characters instead.

5 Results

We conducted the statistical analysis on each trait separately. We used Analysis of Variance (ANOVA) when the normality assumption was not violated (Shapiro–Wilk test for normality) and used the Greenhouse–Geisser degrees of freedom correction when the sphericity assumption was violated (Mauchly's sphericity test). When normality was violated, we used Aligned Rank Transform (ART) instead of ANOVA. For the text-only and appearance experiments (I, II), we had one within-subject factor (Actor). For the unimodal experiments (III, IV), we had two within-subject factors (Voice/Motion, Actor). For the multimodal experiment (V), we had two within-subject factors (Voice, Actor) and one between-subject

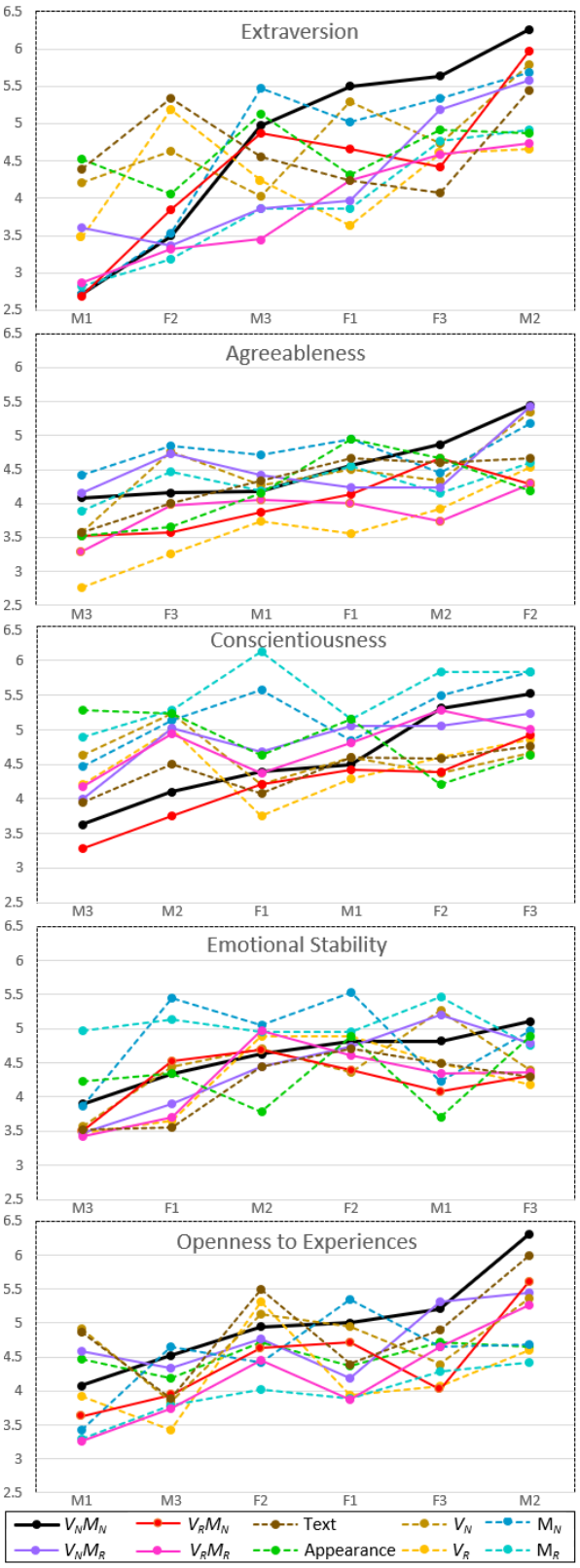

Figure 2: Average trait ratings for each actor. Actor order on the x-axis is based on baseline value (V_NM_N). Shown is a subsection of the full 1-7 scale from 2.5 to 6.5.

Table 3: Mean difference of trait rating for each multimodal condition compared to its unimodal components. E.g. for V_RM_R, we calculate the pairwise difference in rating for each actor to the rating obtained in the voice-only V_R condition (the speech component) and the motion-only M_R condition. The T column represents the text-only condition. Bold text marks the modality with the smallest average pairwise difference.

Condition	Extraversion			Agreeableness			Conscientiousness			Emotional Stability			Openness		
	V	*M*	*T*	*V*	*M*	*T*	*V*	*M*	*T*	*V*	*M*	*T*	*V*	*M*	*T*
V_NM_N	5.16	**1.87**	7.58	**1.87**	2.63	1.97	4.24	3.89	**2.63**	**2.11**	2.97	2.58	3.53	3.84	**3.21**
V_NM_R	4.00	**2.16**	4.95	**1.21**	1.97	2.95	3.03	4.11	**2.58**	1.74	3.74	**1.63**	2.97	4.92	**2.66**
V_RM_N	5.29	**2.53**	4.82	2.76	4.92	**1.89**	3.00	6.37	**2.05**	2.16	3.61	**2.00**	**3.32**	**3.32**	3.74
V_RM_R	3.97	**1.34**	5.87	**2.42**	2.53	2.50	**2.03**	4.55	2.13	**0.82**	4.87	1.13	3.16	**1.74**	4.32

factor (Motion). Post-hoc comparisons were performed with Estimated Marginal Means. Table 4 summarizes all significant results. First, we report results for the appearance control experiment and the text-only experiment. Following this, we structure results based on each personality trait, integrating results from the unimodal Voice and Motion experiments and the multimodal experiment.

5.1 Text-only

For perceptual ratings of text, we found a significant effect of Actor on all personality traits except conscientiousness, mostly in line with our hypothesis. Pairwise comparison of Actor scores across modality conditions revealed that text-only ratings were most similar to those for the V_R condition in three out of five cases (extraversion, conscientiousness, emotional stability), most similar to V_N for openness to experience and M_R for agreeableness.

5.2 Appearance-only

For our static character appearance experiment (II), we only found significant effects on agreeableness and emotional stability. The pairwise differences found for the individual character appearances did not seem to carry over to our motion-only and multimodal experiments (compare Actor effects in Table 4). The differences may be partially due to character gender, supporting our initial hypothesis. When factoring characters by gender, there was a main effect of gender for both agreeableness ($F_{1,18} = 23.70$, $p < .001$, $\eta^2 = 0.57$) and emotional stability ($F_{1,18} = 4.54$, $p < .05$, $\eta^2 = 0.20$), with the male characters perceived as more agreeable ($p < .001$) (post-hoc was non-significant for emotional stability).

5.3 Voice, Motion and Combined Modalities

The following sections (5.3.1 to 5.3.5) detail the results of the Voice-only (III), Motion-only (IV) and Combined Modalities (V) experiments according to each of the Big Five personality traits.

5.3.1 Extraversion

Voice, Motion, and Actor each had a significant effect in the unimodal and multimodal experiments, with Actor identity having the strongest effect. The *Robotic* voice (V_R) was rated as less extraverted, as was the *Robotic* motion (M_R). We also find motion to be particularly important for extraversion, perhaps due to the reduced motion in the robotic condition communicating less nonverbal messages. If an actor is rated as lowly extraverted for speech and highly extraverted for motion, perception of the combined multimodal performance appears to be heavily driven by the motion. That is, actor ratings in the multimodal experiment (V) seem to more closely match ratings in the motion-only experiment (IV) than the voice-only experiment (III). This is evident in Table 3 where the respective unimodal motion condition shows the smaller difference in ratings to the multimodal conditions, and in Figure 2 when comparing the multimodal V_NM_N to the motion-only M_N condition, or the multimodal V_NM_R and V_RM_R to the motion-only M_R condition.

5.3.2 Agreeableness

Voice significantly impacted perceived agreeableness in both the unimodal and the multimodal experiments, whereas Motion only yielded a main effect in the unimodal case. Actor identity showed a main effect for the unimodal Voice and the multimodal experiment, and for both, Voice had the stronger effect. The *Robotic* voice (V_R) was less agreeable and *Robotic* motion (M_R) only less agreeable without the presence of speech. For agreeableness, we find speech to be more important than motion, as is highlighted in Table 3, where the respective unimodal speech conditions yield smaller differences in ratings to the multimodal conditions for all but one case (V_RM_N), in which the smallest difference is that to text. The variety of perceived agreeableness between actors appears to stem from speech, and such variety appears relatively preserved through *Robotic* voice (V_R) processing. Based on motion alone, an individual actor's agreeableness was not distinguished.

5.3.3 Conscientiousness

Voice significantly impacted conscientiousness, but only for multimodal conditions. Motion showed a main effect in both the unimodal and multimodal settings. Actor identity impacted all settings and had a larger effect than Voice or Motion. *Robotic* voice (V_R) was less conscientious than *Natural* voice (V_N), but again, only for multimodal conditions. *Robotic* motion (M_R) was perceived as more conscientious than *Natural* motion (M_N) in both the unimodal and multimodal settings. Perceptions of conscientiousness appear to be formed from the expression of both speech and motion from an actor, where opposing impressions from speech and motion do not override one another. In most cases, we see that perceptions from text-only conditions best mirror those of multimodal, except for V_RM_R, for which voice yields a smaller pairwise difference (Table 3). As such, speech content appears to be the driving factor for perceived conscientiousness, with TTS increasing conscientiousness, possibly due to an association of TTS with informative voice assistants.

5.3.4 Emotional Stability

Voice only had a significant effect on emotional stability in the multimodal setting (V_R was less emotionally stable), whereas Motion did not show a main effect in any setting. Actor identity had the largest effect, with a significant main effect in the unimodal voice and multimodal experiments. For emotional stability, an actor's speech appears to have particular importance, with this information appearing relatively preserved for *Robotic* processing. The driving force of the speech modality can be seen in Table 3, where ratings for the unimodal speech components (voice and text) consistently better mirror the multimodal perceptual rating than the motion component.

5.3.5 Openness to Experience

Voice had a significant effect in both the unimodal and multimodal case. Motion condition only impacted the unimodal setting. Actor identity showed a main effect in all cases. *Robotic* voice (V_R) was rated as less open; *Robotic* motion (M_R) was only perceived as less

open to experience (openness) for the unimodal setting. For openness, there is no clear-cut overriding factor of modality, but rather the combination of voice and motion can yield higher openness than either modality alone. This mixed result can also be seen in Table 3, where no unimodal component consistently outperforms any others.

5.4 Participant Personalities

We conducted a correlation analysis to investigate whether participants' personalities affected their ratings of our characters. For all traits except emotional stability, the self-perceived value correlates with the rating given for that trait (negative correlation coefficient for extraversion, otherwise positive). Positive correlation between self-reported traits and judgements of others is an expected finding, often referred to as "social projection", a common tendency to see similarities between oneself and others [34]. Ratings of emotional stability and openness were additionally influenced by other self-perceived traits. All correlations were minor ($.10 \leq r \leq .17$).

6 Discussion

We investigated how agent personality is transferred or altered across speech and motion modalities according to various degradations. We used motion-capture and speech recordings of six actors with different speaking styles, displaying their performances individually through text, appearance, voice, motion - or a combination of each for a full multimodal representation. We assessed how an actor's performance translates to a virtual agent using Text-to-Speech (TTS) and state-machine-like animation. For the importance of speech and motion modalities, we found that different traits rely differently on each modality. Extraversion appeared to be communicated heavily through motion and was best preserved with high motion fidelity. Previous work also indicates a strong effect of motion appearance on the perception of extraversion [56, 57]. Our findings are consistent with work from Durupinar et al. [17] correlating *'bound flow'* (tense, controlled, contained) with a dampening of perceived extraversion. TTS was perceived as less extraverted, in line with previous findings for effects of pitch [1] and intensity variation [2].

Agreeableness and emotional stability showed the opposite trend; speech drove the perception of a multimodal performance, in line with previous findings [56, 57]. Although TTS (V_R) somewhat retained the original variety between actors, this strongly reduced perceived agreeableness. For conscientiousness and openness to experience, both speech and motion appeared to interplay to form an impression of the character, with conscientiousness showing a trend of favouring the speech modality, specifically the speech content as transcribed text. Work also suggests that conscientiousness and openness are more difficult to judge from short exposure [24, 57].

We assessed how an actor's full performance with voice recordings and full motion-capture ($V_N M_N$) translates to an agent reproducing this performance with TTS and state-machine-like robotic animation ($V_R M_R$). We found that this agent will generally be perceived as less extraverted, agreeable, open and largely less emotionally stable. For conscientiousness, results of the full performance versus the agent were mixed, perhaps due to our finding that TTS decreased conscientiousness, whereas robotic motion increased conscientiousness. The observed dampening effect suggests a need to use actors possessing these traits if the agent is desired to express them. We typically found TTS to decrease ratings of all five personality traits, and robotic motion to decrease extraversion and increase conscientiousness. The effect of robotic motion on conscientiousness indicates that state-machine-like gesture methods may be perceived as more considered and self-disciplined. The consistent dampening effect of the TTS (V_R) may motivate the continuous updating of a virtual agent system to use the best available speech synthesizer, an area of active research [58, 66]. Our results also indicate that an actor's personality plays a role, with lowly extraverted actors being rated as more extraverted and highly extraverted actors as less extraverted when represented by processed or no motion. In the free-form feedback, participants referenced posture and hand gestures as influential on their ratings of motion, with one participant citing both as contributors to their impressions of extraversion, in addition to voice inflection and facial expressions. Participants viewing robotic stimuli appear to rely heavily on verbal features to form impressions, with one participant citing filler words as negatively impacting perceived stability. Natural speech was described as more likeable, warm and calming, while Text-to-Speech was described as emotionless, disconnected, cold and uninteresting.

We selected characters with realistic feature proportions and a cartoon-style to best represent the current fidelity of self-avatars in VR applications (e.g. Meta's Horizon Worlds). Realistic and cartoon avatars are shown to possess similar personality ratings [53], and so, we expect our results to generalize to higher fidelity characters. In our appearance-only experiment, we found only an effect that our male characters were judged as more agreeable. Previous work [71] shows small differences in agreeableness for characters with big differences in appeal, suggesting that our male characters were more appealing than our females or that gender influences agreeableness.

7 Limitations and Future Work

Future work could expose participants to longer stimuli, as our clips were limited by data availability and the need to minimize experiment duration. Datasets containing facial motion capture and personality profiles for each actor would be beneficial. Interaction with agents could facilitate exploring perceptions of user-adaptive agent personalities, including listening and response behavior. Results could support guidelines for 3DUI agent personalities in XR applications. As work finds distance to impact sense of uncanniness and ability to perceive facial expressions [13], immersing participants in VR could impact results through factors other than speech and motion realism. To the best of our knowledge, there is not yet evidence of modality realism impacting perceived personality differently between video- and VR-based settings. VR may increase immersion and presence, potentially augmenting our observations.

Our robotic motion retained the true gestures performed by the actors. However, applications for virtual agents may produce motion using various methods such as automatic generation via machine-learning, rule-based approaches or by applying inverse kinematics algorithms to hand tracking data in VR applications. Future work could investigate application-specific motion processing, in addition to exploring procedural or synthetic motion, as well as multiple levels of motion realism. Finger motion is also desirable for future work, as it may carry significant information about personality [63]. Our findings illustrate how personality could be transferred from multimodal, full performance capture to an agent. Motion appears to be particularly important for portraying extraversion, while speech appears key for agreeableness and emotional stability, with the former heavily affected by the TTS voice.

We propose the following guidelines for designing characters and agents with targeted personalities; (i) use higher fidelity motion to more accurately portray extraversion, (ii) use higher speech realism for portraying higher agreeableness, considering the impact of appearance, (iii) use lower fidelity motion for higher conscientiousness, accounting for interactive effects of speech (iv) focus on designing agent speech for emotional stability and (v) for openness, focus on speech in multimodal contexts, considering motion aspects in unimodal settings as well as a modifying factor.

Acknowledgments

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) (Grant No. 18/CRT/6224), the ADAPT Centre for Digital Content Technology (Grant No. 13/RC/2106_P2) and RADICal (Grant No. 19/FFP/6409).

Table 4: Summary of results: Significant main effects, interactions, and post–hoc analysis. F1-3 denotes our three female actors and M1-3 denotes our three male actors.

Extraversion	**Test**	**Post-hoc**
Text only	ANOVA	
Actor	$F_{5,90} = 5.87$, $p < .001$, $\eta^2 = 0.25$	M2 more extraverted than F1 ($p < .05$) and F3 ($p < .01$). F2 more extraverted than F1 and F3 (both $p < .05$).
Voice only	Aligned Rank Transform	
Voice	$F_{1,198} = 4.90$, $p < .05$, $\eta^2 = 0.22$	V_N more extraverted than V_R ($p < .05$).
Actor	$F_{5,198} = 6.11$, $p < .001$, $\eta^2 = 0.24$	Non-significant
Voice:Actor	$F_{5,198} = 4.05$, $p < .01$, $\eta^2 = 0.24$	For V_N, actor M2 was significantly more extraverted than M1 ($p <$ .0.01) and M3 ($p <$ .0.001). For V_R, actor F2 was significantly more extraverted than F1 ($p < .05$) and M1 ($p <$ .0.01).
Motion only	ANOVA	
Motion	$F_{1,18} = 10.47$, $p <$ .0.01, $\eta^2 = 0.37$	M_N more extraverted than M_R ($p < .001$).
Actor	$F_{5,90} = 17.42$, $p < .001$, $\eta^2 = 0.49$	Actor F2 less extraverted than F1 ($p < .05$), F3 ($p < .001$), M2 ($p < .001$), and M3 ($p < .001$). Actor M1 less extraverted than F1, F3, M2, and M3 (all $p < .001$).
Multimodal	Aligned Rank Transform	
Voice	$F_{1,396} = 7.17$, $p < .01$, $\eta^2 = 0.17$	V_N more extraverted than V_R ($p < .01$)
Motion	$F_{1,36} = 5.36$, $p < .05$, $\eta^2 = 0.11$	M_N more extraverted than M_R ($p < .001$).
Actor	$F_{5,396} = 39.13$, $p < .001$, $\eta^2 = 0.45$	Actor M1 was less extraverted than actors F1, F3, M2, and M3 (all $p < .001$). Actor F2 was also less extraverted than actors F1, F3, M2 (all $p < .001$), and M3 ($p < .05$). Actor M2 was more extraverted than F1 and M3 (both $p < .001$).
Motion:Actor	$F_{5,396} = 4.86$, $p < .001$, $\eta^2 = 0.09$	Appears to follow the directions of the main effect; for all but the one actor rated lowest on this trait (actor M1), *Robotic* Motion visually appears to decrease extraversion (significant only for actor M3).
Agreeableness		
Appearance	ANOVA	
Actor	$F_{5,90} = 4.50$, $p < .01$, $\eta^2 = 0.20$	M1 more agreeable than F1 ($p < .01$) and F2 ($p < .05$). M2 more agreeable than F1 ($p < .05$).
Text only	Aligned Rank Transform	
Actor	$F_{5,90} =$, $p < .05$, $\eta^2 = 0.16$	
Voice only	ANOVA	
Voice	$F_{1,18} = 16.75$, $p < .001$, $\eta^2 = 0.43$	V_N more agreeable than V_R ($p < .001$).
Actor	$F_{5,90} = 10.82$, $p < .001$, $\eta^2 = 0.15$	Actor F2 more agreeable than F1 ($p < .05$), F3 ($p < .01$), M1 ($p < .01$), M2 ($p < .05$), M3 ($p < .001$). Actor M3 additionally less agreeable than F1 ($p < .05$), F3 ($p < .05$), M1 ($p < .05$), M2 ($p < .01$).
Motion only	Aligned Rank Transform	
Motion	$F_{1,198} = 6.80$, $p < .01$, $\eta^2 = 0.25$	M_N more agreeable than M_R ($p < .01$).
Multimodal	Aligned Rank Transform	
Voice	$F_{1,396} = 30.83$, $p < .001$, $\eta^2 = 0.43 = 0.43$	V_N more agreeable than V_R ($p < .001$).
Actor	$F_{5,396} = 6.54$, $p < .001$, $\eta^2 = 0.15 = 0.15$	Actor F2 more agreeable than F1 ($p < .05$), F3 ($p < .01$), M1 ($p < .01$), M3 ($p < .001$). Actor M3 additionally less agreeable than M2 ($p < .05$).
Motion:Actor	$F_{5,396} = 2.50$, $p < .05$, $\eta^2 = 0.06$	Visually, it appears that *Robotic* motion (M_R) decreased agreeableness for some actors (F1, M2), whereas it increased agreeableness for others (actors F3, M1).

Conscientiousness		
Voice only	ANOVA	
Actor	$F_{5,90} = 3.31$, $p < .01$, $\eta^2 = 0.16$	Actor F1 less conscientious than F3 ($p < .05$) and M2 ($p < .001$).
Motion only	Aligned Rank Transform	
Motion	$F_{1,198} = 4.09$, $p < .05$, $\eta^2 = 0.14$	M_N less conscientious than M_R ($p < .05$).
Actor	$F_{5,198} = 8.89$, $p < .001$, $\eta^2 = 0.36$	Actor M3 less conscientious than F1 ($p < .001$), F2 ($p < .01$), F3 ($p < .001$).
Multimodal	Aligned Rank Transform	
Voice	$F_{1,396} = 3.94$, $p < .05$, $\eta^2 = 0.13$	V_R less conscientious ($p < .05$).
Motion	$F_{1,36} = 4.27$, $p < .05$, $\eta^2 = 0.09$	M_N less conscientious than M_R ($p < .05$).
Actor	$F_{5,396} = 11.55$, $p < .001$, $\eta^2 = 0.19$	Actor M3 less conscientious than F2 ($p < .001$), F3 ($p < .001$), M1 ($p < .001$), M2 ($p < .05$). Actor F3 additionally more conscientious than F1 ($p < .05$) and M2 ($p < .05$).
Emotional Stability		
Appearance	ANOVA	
Actor	$F_{5,90} = 3.77$, $p < .01$, $\eta^2 = 0.17$	M3 more emotionally stable than M2 ($p < .05$)
Text only	ANOVA	
Actor	$F_{5,90} = 2.86$, $p < .05$, $\eta^2 = 0.14$	Non-significant
Voice only	Aligned Rank Transform	
Actor	$F_{5,198} = 2.25$, $p < .001$, $\eta^2 = 0.26$	Actor F1 less emotionally stable than actors F2 ($p < .05$), M1 ($p < .01$), M2 ($p < .01$). Actor M3 less emotionally stable than actors F2 ($p < .001$), F3 ($p < .01$), M1 ($p < .001$), M2 ($p < .001$). Actor M1 more emotionally stable than actor F3 ($p < .05$).
Motion only	Aligned Rank Transform	
Actor (marginal)	$p = 0.55$	Non-significant
Motion:Actor	$F_{5,198} = 3.46$, $p < .01$, $\eta^2 = 0.16$	Visually, it appears that only the two actors rated lowest on emotional stability for M_N (M1 & M3), were perceived more emotionally stable under M_R.
Multimodal	Aligned Rank Transform	
Voice	$F_{1,396} = 5.10$, $p < .05$, $\eta^2 = 0.11$	V_R less emotionally stable ($p < .05$).
Actor	$F_{5,396} = 9.16$, $p < .001$, $\eta^2 = 0.16$	Actor F1 less emotionally stable than actors F2, F3, M1 (all $p < .05$), and M2 ($p < .01$). Actor M3 less emotionally stable than actors F1 ($p < .05$), F2, F3, M1, and M2 (all $p < .001$).
Openness to Experience		
Text only	ANOVA	
Actor	$F_{5,90} = 10.91$, $p < .001$, $\eta^2 = 0.38$	F2 more open than F1 ($p < .05$). M2 more open than F1 ($p < .001$), F3 ($p < .05$), M1 ($p < .05$), and M3 ($p < .001$). M3 less open than F2 ($p < .001$).
Voice only	ANOVA	
Voice	$F_{1,18} = 10.57$, $p < .01$, $\eta^2 = 0.37$	V_N more open ($p < .001$).
Actor	$F_{3.7,67.5} = 7.14$, $p < .001$, $\eta^2 = 0.28$	Actor M3 less open than F1 ($p < .05$), F2 ($p < .001$), M2 ($p < .001$). Actor F2 additionally more open than F3 ($p < .01$).
Voice:Actor	$F_{5,90} = 2.35$, $p < .05$, $\eta^2 = 0.12$	For most actors, V_R appeared to decrease openness to experience, but this was reversed for actor F2 (no relevant pairwise comparisons significant)
Motion only	Aligned Rank Transform	
Motion	$F_{1,198} = 10.75$, $p < .01$, $\eta^2 = 0.34$	M_N more open than M_R ($p < .01$).
Actor	$F_{5,198} = 4.11$, $p < .01$, $\eta^2 = 0.15$	Actor M1 less open than F1, F3, M2 (all $p < .01$).
Multimodal	ANOVA	
Voice	$F_{1,36} = 14.64$, $p < .001$, $\eta^2 = 0.29$	V_N more open ($p < .001$).
Actor	$F_{5,180} = 14.38$, $p < .001$, $\eta^2 = 0.29$	Actor M2 more open than F1 ($p < .001$), F2 ($p < .001$), F3 ($p < .01$), M1 ($p < .001$), M3 ($p < .001$). Actor M1 also less open than F2 and F3 (both $p < .01$).

REFERENCES

[1] D. W. Addington. The relationship of selected vocal characteristics to personality perception. 1968.

[2] C. D. Aronovitch. The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of social psychology*, 99(2):207–220, 1976.

[3] M. Astrid, N. C. Krämer, and J. Gratch. How our personality shapes our interactions with virtual characters - implications for research and development. In *International Conference on Intelligent Virtual Agents*, pp. 208–221. Springer, 2010.

[4] M. P. Aylett, A. Vinciarelli, and M. Wester. Speech synthesis for the generation of artificial personality. *IEEE Transactions on Affective Computing*, 11(2):361–372, 2020. doi: 10.1109/TAFFC.2017.2763134

[5] G. Bente, F. Eschenburg, and L. Aelker. Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. 2007.

[6] D. A. Bray and B. Konsynski. Virtual worlds, virtual economies, virtual institutions. In *Virtual Worlds and New Realities Conference at Emory University*, 2008.

[7] J. Brixey and D. Novick. Building rapport with extraverted and introverted agents. In *Advanced Social Interaction with Agents*, pp. 3–13. Springer, 2019.

[8] S. Burke, T. Bresnahan, T. Li, K. Epnere, A. Rizzo, M. Partin, R. Ahlness, and M. Trimmer. Using virtual interactive training agents (vita) with adults with autism and other developmental disabilities. *Journal of Autism and Developmental Disorders*, 48, 03 2018. doi: 10.1007/s10803-017-3374-z

[9] J. P. Cabral, B. R. Cowan, K. Zibrek, and R. McDonnell. The influence of synthetic voice on the evaluation of a virtual character. In *INTERSPEECH*, pp. 229–233, 2017.

[10] A. Cafaro, H. H. Vilhjálmsson, T. Bickmore, D. Heylen, K. R. Jóhannsdóttir, and G. S. Valgarðsson. First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In *International conference on intelligent virtual agents*, pp. 67–80. Springer, 2012.

[11] D. Chi, M. Costa, L. Zhao, and N. Badler. The emote model for effort and shape. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, p. 173–182. ACM Press/Addison-Wesley Publishing Co., USA, 2000. doi: 10.1145/344779.352172

[12] E. K. Chiou, N. L. Schroeder, and S. D. Craig. How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education*, 146:103756, 2020.

[13] Z. Choudhary, K. Kim, R. Schubert, G. Bruder, and G. F. Welch. Virtual big heads: Analysis of human perception and comfort of head scales in social virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 425–433, 2020. doi: 10.1109/VR46266.2020.00063

[14] A. W. de Borst and B. de Gelder. Is it the real deal? perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Frontiers in Psychology*, 6:576, 2015. doi: 10.3389/fpsyg.2015.00576

[15] B. De Carolis, C. Pelachaud, I. Poggi, and M. Steedman. Apml, a markup language for believable behavior generation. In *Life-like characters*, pp. 65–85. Springer, 2004.

[16] J. D. N. Dionisio, W. G. B. III, and R. Gilbert. 3d virtual worlds and the metaverse: Current status and future possibilities. *ACM Comput. Surv.*, 45(3), July 2013. doi: 10.1145/2480741.2480751

[17] F. Durupinar, M. Kapadia, S. Deutsch, M. Neff, and N. I. Badler. Perform: Perceptual approach for adding ocean personality to human motion using laban movement analysis. *ACM Trans. Graph.*, 36(1), Oct. 2016. doi: 10.1145/2983620

[18] J. Ehret, A. Bönsch, L. Aspöck, C. T. Röhr, S. Baumann, M. Grice, J. Fels, and T. W. Kuhlen. Do prosody and embodiment influence the perceived naturalness of conversational agents' speech? *ACM Trans. Appl. Percept.*, 18(4), oct 2021. doi: 10.1145/3486580

[19] C. Ennis, L. Hoyet, A. Egges, and R. McDonnell. Emotion capture: Emotionally expressive characters for games. In *Proceedings of Motion on Games*, MIG '13, p. 53–60. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2522628.2522633

[20] Y. Ferstl, M. Neff, and R. McDonnell. Adversarial gesture generation with realistic gesture phasing. *Computers & Graphics*, 89:117–130, 2020.

[21] Y. Ferstl, M. Neff, and R. McDonnell. Expressgesture: Expressive gesture generation from speech through database matching. *Computer Animation and Virtual Worlds*, p. e2016, 2021.

[22] Y. Ferstl, S. Thomas, C. Guiard, C. Ennis, and R. McDonnell. Human or robot? investigating voice, appearance and gesture motion realism of conversational social agents. In *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents*, IVA '21, p. 76–83. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472306.3478338

[23] A. Gluszek and J. F. Dovidio. The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and social psychology review*, 14(2):214–237, 2010.

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

[25] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.

[26] K. ISBISTER and C. NASS. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, 53(2):251–267, 2000. doi: 10.1006/ijhc.2000.0368

[27] M. Jensen. Personality traits and nonverbal communication patterns. *International Journal of Social Science Studies*, 4:57–70, 03 2016. doi: 10.11114/ijsss.v4i5.1451

[28] O. P. John, E. M. Donahue, and R. L. Kentle. Big five inventory. *Journal of Personality and Social Psychology*, 1991.

[29] M. Kipp, M. Neff, and I. Albrecht. An annotation scheme for conversational gestures: how to economically capture timing and form. *Language Resources and Evaluation*, 41(3):325–339, 2007.

[30] M. Kipp, M. Neff, K. H. Kipp, and I. Albrecht. Towards natural gesture synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In *International workshop on intelligent virtual agents*, pp. 15–28. Springer, 2007.

[31] S. Kita, I. Van Gijn, and H. Van der Hulst. Movement phases in signs and co-speech gestures, and their transcription by human coders. In *International Gesture Workshop*, pp. 23–35. Springer, 1997.

[32] M. Koppensteiner, P. Stephan, and J. P. M. Jäschke. More than words: Judgments of politicians and the role of different communication channels. *Journal of Research in personality*, 58:21–30, 2015.

[33] R. S. S. Kramer and R. Ward. Internal facial features are signals of personality and health. *The Quarterly Journal of Experimental Psychology*, 63(11):2273–2287, 2010. doi: 10.1080/17470211003770912

[34] J. Krueger. On the perception of social consensus. *Advances in experimental social psychology*, 30:163–240, 1998.

[35] A. Kuzminykh, J. Sun, N. Govindaraju, J. Avery, and E. Lank. Genie in the bottle: Anthropomorphized perceptions of conversational agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

[36] G. Lee, Z. Deng, S. Ma, T. Shiratori, S. Srinivasa, and Y. Sheikh. Talking with hands 16.2m: A large-scale dataset of synchronized body-finger motion and audio for conversational motion analysis and synthesis. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 763–772, 2019. doi: 10.1109/ICCV.2019.00085

[37] J. Lee and S. Marsella. Nonverbal behavior generator for embodied conversational agents. In *International Workshop on Intelligent Virtual Agents*, pp. 243–255. Springer, 2006.

[38] K. Legde and D. W. Cunningham. Evaluating the effect of clothing and environment on the perceived personality of virtual avatars. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, IVA '19, p. 206–208. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3308532.3329425

[39] M. Levesque and D. Kenny. Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology*, 65:1178–1187, 12 1993. doi: 10.1037/0022-3514.65.6.1178

[40] C. F. Lima, A. Anikin, A. C. Monteiro, S. K. Scott, and S. L. Castro. Automaticity in the recognition of nonverbal emotional vocalizations. *Emotion*, 19(2):219, 2019.

[41] A. C. Little and D. I. Perrett. Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology*, 98(1):111–126, 2007. doi: 10.1348/000712606X109648

[42] F. Mairesse and M. Walker. Personage: Personality generation for dialogue. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pp. 496–503, 2007.

[43] P. McAleer, A. Todorov, and P. Belin. How do you say 'hello'? personality impressions from brief novel voices. *PLOS ONE*, 9(3):1–9, 03 2014. doi: 10.1371/journal.pone.0090779

[44] M. B. Moussa, Z. Kasap, N. Magnenat-Thalmann, K. Chandramouli, S. N. Haji Mirza, Q. Zhang, E. Izquierdo, I. Biperis, and P. Daras. Towards an expressive virtual tutor: An implementation of a virtual tutor based on an empirical study of non-verbal behaviour. In *Proceedings of the 2010 ACM Workshop on Surreal Media and Virtual Cloning*, SMVC '10, p. 39–44. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1878083.1878096

[45] C. Nass and K. M. Lee. Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, 7(3):171–181, 2001.

[46] L. P. Naumann, S. Vazire, P. J. Rentfrow, and S. D. Gosling. Personality judgments based on physical appearance. *Personality and social psychology bulletin*, 35(12):1661–1671, 2009.

[47] M. Neff, Y. Wang, R. Abbott, and M. Walker. Evaluating the effect of gesture and language on personality perception in conversational agents. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, eds., *Intelligent Virtual Agents*, pp. 222–235. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[48] J. Ondřej, C. Ennis, N. A. Merriman, and C. O'sullivan. Frankenfolk: Distinctiveness and attractiveness of voice and motion. *ACM Trans. Appl. Percept.*, 13(4), July 2016. doi: 10.1145/2948066

[49] X. Pan, M. Gillies, and M. Slater. Virtual character personality influences participant attitudes and behavior – an interview with a virtual human character about her social anxiety. *Frontiers in Robotics and AI*, 2:1, 2015. doi: 10.3389/frobt.2015.00001

[50] J. W. Pennebaker and L. A. King. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296, 1999.

[51] M. Perez Garcia and S. Saffon Lopez. *Exploring the Uncanny Valley Theory in the Constructs of a Virtual Assistant Personality*, pp. 1017–1033. 01 2020. doi: 10.1007/978-3-030-29516-5_76

[52] T. Polzehl. *PERSONALITY IN SPEECH*. Springer, 2016.

[53] K. Ruhland, K. Zibrek, and R. McDonnell. Perception of personality through eye gaze of realistic and cartoon models. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, pp. 19–23, 2015.

[54] A. Shapiro, P. Faloutsos, and V. Ng-Thow-Hing. Dynamic animation and control environment. In *Proceedings of graphics interface 2005*, pp. 61–70. Canadian Human-Computer Communications Society, 2005.

[55] M. Shvo, J. Buhmann, and M. Kapadia. An interdependent model of personality, motivation, emotion, and mood for intelligent virtual agents. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, IVA '19, p. 65–72. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3308532.3329474

[56] H. J. Smith and M. Neff. Understanding the impact of animated gesture performance on personality perceptions. *ACM Trans. Graph.*, 36(4), July 2017. doi: 10.1145/3072959.3073697

[57] S. Sonlu, U. Güdükbay, and F. Durupinar. A conversational agent framework with multi-modal personality expression. *ACM Trans. Graph.*, 40(1), Jan. 2021. doi: 10.1145/3439795

[58] É. Székely, G. E. Henter, J. Beskow, and J. Gustafson. Spontaneous conversational speech synthesis from found data. In *INTERSPEECH*, pp. 4435–4439, 2019.

[59] M. T. Tang, V. L. Zhu, and V. Popescu. Alterecho: Loose avatar-streamer coupling for expressive vtubing. 2021.

[60] M. Thiebaux, S. Marsella, A. N. Marshall, and M. Kallmann. Smartbody: Behavior realization for embodied conversational agents. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, pp. 151–158, 2008.

[61] I. Torre, E. Carrigan, R. McDonnell, K. Domijan, K. McCabe, and N. Harte. The effect of multimodal emotional expression and agent appearance on trust in human-agent interaction. In *Motion, Interaction and Games*, MIG '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359566.3360065

[62] A. Vinciarelli, M. Pantic, H. Bourlard, and A. Pentland. Social signal processing: State-of-the-art and future perspectives of an emerging domain. In *Proceedings of the 16th ACM International Conference on Multimedia*, MM '08, p. 1061–1070. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1459359.1459573

[63] Y. Wang, J. E. F. Tree, M. Walker, and M. Neff. Assessing the impact of hand motion on virtual character personality. *ACM Transactions on Applied Perception (TAP)*, 13(2):1–23, 2016.

[64] J. Willis and A. Todorov. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science*, 17(7):592–598, 2006.

[65] K. Wolffhechel, J. Fagertun, U. P. Jacobsen, W. Majewski, A. S. Hemmingsen, C. L. Larsen, S. K. Lorentzen, and H. Jarmer. Interpretation of appearance: The effect of facial features on first impressions and personality. *PLOS ONE*, 9(9):1–8, 09 2014. doi: 10.1371/journal.pone.0107721

[66] Y. Yamashita, T. Koriyama, Y. Saito, S. Takamichi, Y. Ijima, R. Masumura, and H. Saruwatari. Investigating effective additional contextual factors in dnn-based spontaneous speech synthesis. In *INTERSPEECH*, pp. 3201–3205, 2020.

[67] E. Zell, K. Zibrek, and R. McDonnell. Perception of virtual characters. In *ACM SIGGRAPH 2019 Courses*, SIGGRAPH '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3305366.3328101

[68] E. Zell, K. Zibrek, X. Pan, M. Gillies, and R. McDonnell. From perception to interaction with virtual characters. 2020.

[69] M. X. Zhou, G. Mark, J. Li, and H. Yang. Trusting virtual agents: The effect of personality. *ACM Trans. Interact. Intell. Syst.*, 9(2–3), Mar. 2019. doi: 10.1145/3232077

[70] K. Zibrek, E. Kokkinara, and R. Mcdonnell. The effect of realistic appearance of virtual characters in immersive environments - does the character's personality play a role? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1681–1690, 2018. doi: 10.1109/TVCG.2018.2794638

[71] K. Zibrek and R. McDonnell. Does render style affect perception of personality in virtual humans? In *Proceedings of the ACM Symposium on Applied Perception*, SAP '14, p. 111–115. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2628257.2628270