How Do You Like Your Virtual Agent?: Human-Agent Interaction Experience through Nonverbal Features and Personality Traits

Aleksandra Cerekovic[1,2], Oya Aran[1], and Daniel Gatica-Perez[1,3]

[1] Idiap Research Institute, Martigny, Switzerland
[2] University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
[3] Ecole Polytechnique Federal de Lausanne (EPFL), Lausanne, Switzerland

Abstract. Recent studies suggest that human interaction experience with virtual agents can be, to a very large degree, described by people's personality traits. Moreover, the nonverbal behavior of a person has been known to indicate several social constructs in different settings. In this study, we analyze human-agent interaction from the perspective of the personality of the human and the nonverbal behaviors he/she displays during the interaction. Based on existing work in psychology, we designed and recorded an experiment on human-agent interactions, in which a human communicates with two different virtual agents. Human-agent interactions are described with three self-reported measures: quality, rapport and likeness of the agent. We investigate the use of self-reported personality traits and extracted audio-visual nonverbal features as descriptors of these measures. Our results on a correlation analysis show significant correlations between the interaction measures and several of the personality traits and nonverbal features, which are supported by both psychology and human-agent interaction literature. We further use traits and nonverbal cues as features to build regression models for predicting measures of interaction experience. Our results show that the best results are obtained when nonverbal cues and personality traits are used together.

Keywords: human-agent interaction, quality of interaction, nonverbal behavior, Big 5 personality traits.

1 Introduction

A growing number of applications seek to provide social abilities and human-like intelligence to computers. Compelling social interactions with computers, or specifically Embodied Conversational Agents (ECAs), are persuasive, engaging, and they increase trust and feeling of likeness, so it is understandable why recent trends show increasing usage of virtual agents in social media, education or social coaching.

Clearly, with the advance of social, user-aware adaptive interfaces, it has become increasingly important to model and reason social judgment for agents.

H.S. Park et al. (Eds.): HBU 2014, LNCS 8749, pp. 1–15, 2014.

To help virtual agents to interpret the human behaviors a number of observation studies has been proposed: human conversational behaviors are induced in (mainly, with the Wizard-of-Oz) experiments with agents, or in interaction with other humans. Further, observed behaviors are used to model both perception and reasoning components for the agents.

Several studies investigated the impact of human personality on the outcomes of human-agent interaction (HAI) and on the evaluation of the agent, with a goal to understand human preference for interactive characters. In most of those works, only the extraversion trait has been considered as the personality trait to analyze. The most notable studies on this topic come from the early 2000s. Limited by technology, researchers used only vocal behaviors [19], or a still image and textual interfaces [10] to simulate the extraverted/intraverted agent. In similar and more recent studies extraversion is manipulated via computer-generated voice and gestures of 2D cartoon-like agent [5]. As outcomes, it has been shown how humans are attracted by characters who have both similar personality, confirming similarity rule, and opposite personality, confirming complementary rule (see [5] for an overview).

Other recent studies have started to observe influence of personality traits other than extraversion to various social phenomena of HAI, such as rapport, or perception of agent's personality [17]. In [28], two conditions (low and high behaviour realism) of an agent designed to build rapport (the Rapport agent) were manipulated in interaction with humans. Further, human personality traits were correlated with persistent behavioral patterns, such as shyness or fear of interpersonal encounters. The results of the study have shown how both extraversion and agreeableness have been recognized to have a major impact on human attitudes, more than gender and age. Other Big 5 traits, namely neuroticism, openness to experience and consciousness were not found significant. Another study with the Rapport agent compared the perceived rapport of HAI to the rapport experienced in human-human conversation [12]. Results indicate how people who score higher in agreeableness perceived strong rapport both with the agent and a human, with a stronger relationship for the agent than human. Moreover, people with higher conscientiousness reported strong rapport when they communicated with both the agent and a human. A first-impression study [6], analyzed the impact of human personality on human judgments of the agents across conditions in which agents displayed different nonverbal behaviors (proximity and amount of smiles and gazing). Judgments included agent's extraversion and friendliness. The study has shown how agent smiles had a main effect on judging of friendliness, showing positive correlation between smiles and friendliness. However, the relation between human personality and perceived interaction in this study is not that evident: it has only been concluded that people with low agreeableness tend to interpret agents who gaze more as friendlier.

In this paper, we build an experimental study to investigate the influence of human personality to perceived experience of HAI. We also study how humans' audio-visual nonverbal cues can be used to reveal perceived experience. We further experiment with regression models to predict the perceived experience

measures using both personality traits and nonverbal cues. The motivation for our study comes from several facts. As explained beforehand, personality traits shape perception and behaviors of humans in human-human and human-agent interaction. Nonverbal cues have been also shown to characterize several social constructs [13], and to be significant in predicting some of the Big 5 traits in social computing (e.g. in social media [3], and in face-to-face meetings [1]). Moreover, recent advances in social computing have shown how fusion of audio-visual data is significant for prediction of various behavioral patterns and phenomena in social dialogue, such as dominance [25] or aggression [14]. Thus, we believe that fusion of both visual and acoustic cues could be significant for predicting perceived measures of HAI. Our study is similar to the study with the Rapport agent [28], but with one major difference: rather than only observing the influence of personality traits on HAI experience we focus on the multi-modal analysis of perceived experience using both visual and vocal nonverbal behavior cues as well as the personality traits.

Specifically, in this paper we investigate the nonverbal cues and self-reported Big five traits as descriptors of an interaction of a person with two different virtual agents. We design a study in which we collect audio-visual data of humans talking with agents, along with their Big 5 traits and perceived experience measures. We describe interaction experience through three measures (quality, rapport and likeness) [7]. The virtual agents we use in our study are Sensitive Artificial Listeners (SALs)[17], which are designed with the purpose of inducing specific emotional conversation. There are in total four different agents in the SAL system: happy, angry, sad and neutral character. Studies suggest that the perceived personality of a social artifact has a significant effect on usability and acceptance [27], so we find these agents relevant to explore the interaction experience. Though SALs' understanding capabilities are limited to emotional processing, their personality has been successfully recognized in a recent evaluation study [17].

Our study has three contributions. First, we examine the relation between the self-reported Big 5 traits and perceived experience in human-agent interaction, with comparison to existing work in social psychology and human-agent interaction. Second, we investigate links between nonverbal cues and perceived experience, with an aim to find which nonverbal patterns are significant descriptors of experience aspects: quality of interaction, rapport and likeness of the agent. Finally, we build a method to predict HAI experience outcome based on automatically extracted nonverbal cues displayed during the interaction and self-reported Big 5 traits. Given the fact that we record our subjects with a consumer depth camera, we also investigate and discuss potentials of using cheap markerless tracking system for analyzing nonverbal behaviors.

2 Data Collection

Our data collection contains recordings of 33 subjects, out of which are 14 females and 19 males. 26 are graduate students and researchers in computer science, and

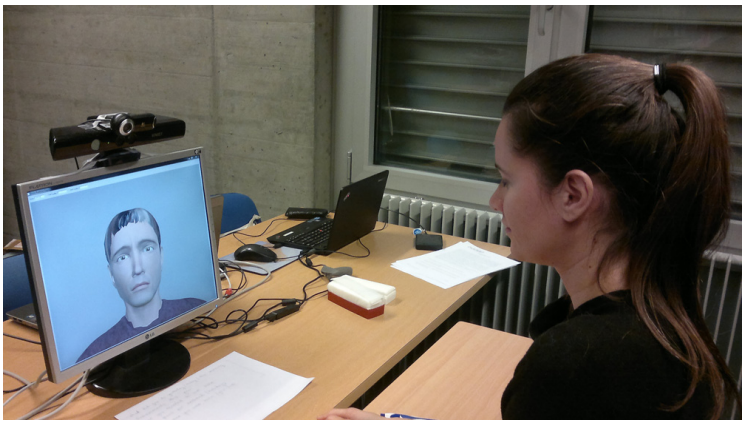

Fig. 1. Recording environment with a participant

7 are students of management. Most of them have different cultural background; however 85% subjects are Caucasians. Subjects were recruited using two mailing lists and they were compensated with 10 CHF for participation.

Before the recording session, each subject had to sign the consent form, and fill in demographic information and NEO FFI Big 5 personality questionnaire [16]. The recording session contains three recordings of the subject, where the data has been captured with a Kinect RGB-D camera (see Figure 1). First, the subject was asked to give a 1-minute self-presentation via video call. Then, he/she had two 4-minute interactions with two agents: first interaction was with sad Obadiah, and second with cheerful Poppy. These characters are selected because evaluation study on SALs [17] has shown how Poppy is the most consistent and familiar and Obadiah is the most believable character. Before the interaction, subjects were given an explanation what SALs are and what they can expect from interaction. To encourage the interaction, a list of potential conversation topics was placed in the view-field of a subject. Topics were: plans for the weekend, vacation plans, things that a subject did yesterday/last weekend, country where a subject was born, last book which a subject read. After each human-agent interaction, the subjects filled out a questionnaire, reporting their perceived interaction experience and mood. Due to the relatively small number of recruited subjects, we assigned all subjects to same experimental conditions, meaning that they first interacted with sad Obadiah, then to cheerful Poppy.

Interaction experience measures have been inspired from the study [7] in which authors investigate how Big 5 traits are manifested in mixed-sex dyadic interactions of strangers. To measure perceived interaction, they construct a "Perception of Interaction" questionnaire with items which rate various aspects of participants' interaction experience. We target the same aspects in human-agent interaction: Quality of Interaction (QoI), Degree of Rapport (DoR) and Degree of Likeness of the agent (DoL). Each interaction aspect in our questionnaire was targeted by a group of statements with a five-point Likert scale ((1) - Disagree strongly to (5) - Agree strongly).

Some of the items used by [7] were excluded, such as "I believe that partner wants to interact more in the future", given the constrained social and perception abilities of SALs. In total, our interaction questionnaire has 15 items which report QoI (7), DoR (5) and DoL (3). The questions that we used in the questionnaire and the target aspect of each question is shown in Table 1. The values of these measures are normalized to the range in [0, 1]. Additionally, our questionnaire also measures subject's mood (same questionnaire as used in [4]), which is at the moment excluded from our experiments.

Table 1. The questions and targeted aspects in the interaction questionnaire

Question	Target Aspect
The interaction with the character was smooth, natural, and relaxed.	QoI
I felt accepted and respected by the character.	DoR
I think the character is likable.	DoL
I enjoyed the interaction	QoI.
I got along with the character pretty good.	DoR
The interaction with the character was forced, awkward, and strained.	QoI
I did not want to get along with the character.	DoL
I was paying attention to way that character responds to me and I was adapting my own behaviour to it.	DoR
I felt uncomfortable during the interaction.	QoI
The character often said things completely out of place.	QoI
I think that the character finds me likable.	DoR
The interaction with the character was pleasant and interesting.	QoI
I would like to interact more with the character in the future.	DoL
I felt that character was paying attention to my mood.	DoR
I felt self-conscious during the conversation.	QoI

At the end of each recording session, several streams were obtained: RGB-D data and audio data from Kinect, and screen captures and log files with description of agent's behaviour.

3 Cue Extraction

We extracted nonverbal cues from both visual and auditory channel. The selection of features was based on previous studies on human-human interaction and conversational displays in psychology. For visual nonverbal displays we studied the literature on displays of attitude in initial human-human interactions (interactions where the interaction partners meet for the first time). Then, given the fact that previous research has shown how personality traits of extraversion and agreableness are important predictors of HAI [28], we also take into account findings on nonverbal cues which are important for predicting personality traits.

Related to attitude in initial human-human interaction, a number of works observe how postural congruence and mimicry are positively related to liking and rapport ([15,31], or more recently [29]). Mimicry has also been investigated in human-agent community, with attempts to build automatic models to predict mimicry [26]. Our interaction scenario can only observe facial mimicry, because SAL agents have only their face visible and they do not make any body leans. Among other nonverbal cues, psychological literature agrees how frequent eye contact, relaxation, leaning and orienting towards, less fiddling, moving closer, touching, more open arm and leg positions, smiling and more expressive face and voice are signs of liking from observer's (or coder's) point of view [2,13]. Yet, when it comes to displays of liking associated with self-reported measures, findings are not that evident. In an extensive review of literature dealing with the posture cue, Mehrabian shows how displays of liking vary from gender and status [18]. He also shows how larger reclining angle of sideways leaning communicates a more negative attitude, and smaller reclining angle of a communicator while seated, and therefore a smaller degree of trunk relaxation, communicates a more positive attitude. Investigation of non-verbal behavior cues and liking conducted on initial same-sex dyad interactions [15] shows how the most significant variables in predicting subjects' liking is the actual amount of mutual gaze and the total percentage time looking. Other significant behaviors are: expressiveness of the face and the amount of activity in movement and gesture, synchrony of movement and speech, and expressiveness of the face and gesturing. Another cross-study [24] examined only kinesics and vocalic behaviors. Results show how increased pitch variety is associated with female actors, whereas interesting effect is noticed for loudness and length of talking, which decrease over interaction time. Though authors say how their research shows how this means disengagement in conversations, another work reports how this means greater attractiveness [21].

Psychologists have noted that, when observed alone, vocal and paralinguistic features have the highest correlation with person judgments of personality traits, at least in certain experimental conditions [8]. This has been confirmed in some studies in automatic recognition of personality traits which use nonverbal behavior as predictors. A study on the prediction of personality impressions analyses predictability of Big 5 personality trait impressions using audio-visual nonverbal cues extracted from the vlogs [3]. Nonverbal cues include speaking activity (speaking time, pauses, etc.), prosody (spectral entropy, pitch, etc.), motion (weigthed motion energy images, movements in front of camera), gaze behavior, vertical framing (position of the face), and distance to camera. Among the cues, speaking time and length, prosody, motion and looking time were most significant for inferring the perceived personality. Observer judgments of extraversion are positively correlated with high fluency, meaning greater length of the speech segments, and less number of speaking turns, and positively with loudness, looking time and motion. People who are observed as more agreeable speak with higher voice, and people who are observed as more extraverted have a higher vocal control. In another study on meeting videos [23], speech related measurements (e.g., speaking

time, mean energy, pitch, etc.) and percent of looking time (e.g., amount of received and given gaze) were shown as significant predictors of personality traits.

Based on the overviewed literature we extract the following features from human-agent interaction sequences: speaking activity, prosody, body leans, head direction, visual activity, and hand activity. Every cue, except hand activity, is extracted automatically from whole conversational sequences. Whereas we acknowledge the importance of mimicry, in this experiment we only extract individual behaviors of humans without looking at agent's behavior.

3.1 Audio Cues

To extract nonverbal cues from speech, we first applied automatic speaker diarization on human-agent audio files using Idiap Speaker Diarization Toolkit [30]. We further used MIT Human Dynamics group toolkit ([22] to export voice quality measures.

Speaking Activity. Based on the diarization output, we extracted the speech segments of the subject and computed the following features for each human-agent sequence: total speaking length (TSL), total speaking turns (TST), filtered turns, and average turn duration (ATD).

Voice Quality Measures. The voice quality measures are extracted on the subject's speech, based on the diarization output. We extracted the statistics - mean and standard deviation - of following features: pitch (F0 (m), F0 (std)), pitch confidence (F0 conf (m), F0 conf (std)), spectral entropy (SE (m), SE (std)), delta energy (DE (m), DE (std)), location of autocorrelation peaks (Loc R0 (m), Loc R0 (std), number of autocorrelation peaks (# R0 (m), # R0 (std)), value of of autocorrelation peaks (Val R0 (m), Val R0 (std)). Furthermore, three other measures were exported: average length of speaking segment (ALSS), average length of voiced segment (ALVS), fraction of time speaking (FTS), voicing rate (VR), and fraction speaking over (FSO).

3.2 Visual Cues

One of the aspects we wanted to investigate in this study is the potential of using cheap markerless motion capture systems (MS Kinect SDK v1.8) for the purpose of automatic social behavior analysis. Using Kinect SDK upper body and face tracking information we created body lean and head direction classifier. Since the tracker produced significantly poor results for arm/hand joints, hand activity of the subject during the interaction was manually annotated.

Body Leans. In this paper we propose a module for automatic analysis of body leans from 3D upper body pose and depth image. We use a support vector machine (SVM) classifier, RBF kernel, trained with extended 3D upper body pose features. Extended 3D upper body pose is an extended version of features extracted from Kinect SDK upper body tracker; along with x-y position values of shoulders, neck and head, it also contains torso information and z-values of

shoulders, neck and torso normalized with respect to the neutral body pose. Using our classifier, distribution of the following body leans is extracted: neutral, sideways left, sideways right (SR), forward and backward leans (BL). These categories are inspired from psychological work on posture behavior and displays of affect [18]. Along with those distributions we also compute frequency of shifting between those leans.

Head Direction. We use a simple method which outputs three head directions; screen, table, or other (HDO), and frequency of shifts (HDFS). The method is using 3D object approximation of screen and table and head information retrieved from Kinect face tracker. The method is tested on manually annotated ground truth data and is proven to produce satisfying results.

Visual Activity. The visual activity of the subject is extracted by using weighted motion energy images (wMEI), which is a binary image that describes the spatial motion distribution in the video sequence [3]. The features we extract are statistics of wMEI: entropy, mean and median value.

Hand Activity. To manually annotate hand activity we used the following classes: hidden hands, hand gestures (GES), gestures on table, hands on table, and self-touch (ST). The classes are proposed in a study on body expressions of participants of employment interviews [20].

4 Analysis and Results

In the first two parts of this section, we present the correlation analysis and links between the interaction experience and Big 5 traits and also the extracted nonverbal cues. We compare and discuss our results with previous works from psychology and human-agent interaction literature. We also present the results of our experiments for predicting interaction experience.

4.1 Personality and Interaction Experience

We find the individual correlations between Big 5 traits of the participants and individual measures of interaction experience to understand what traits may be useful to infer interaction with two virtual characters.

Table 2 shows the significant correlations. Extraversion has the highest correlations with both agents; it is then followed by neuroticism and agreeableness. With regard to extraversion, we found that extraverted subjects reported good QoI and high DoR to both of agents. Extraverted people also reported high DoL for Obadiah, whereas for Poppy we found no significant evidence. In a study on human-human interaction which inspired our work ([7]) the extraverted people were more likely to report that they did not feel self-conscious, they perceived their interaction to be smooth, natural, and relaxed, and they felt comfortable around their interaction partner. The similar study on Big Five manifestation in initial dyadic interactions[9] has also shown how extraverted people tend to rate interaction natural and relaxed. This is a direct reference to Carl Jungs view

Table 2. Significant Pearson correlation effects between Big Five traits and interaction experience measures: QoI, DoR and DoL ($p < 0.05$,*$p < 0.01$)

	Obadiah	Poppy
Openness to Exp.	-	-
Conscientiousness	QoI (.41)	-
Extraversion	QoI (.44) DoR (.58)* DoL (.42)	QoI (.36) DoR (.44)
Agreeableness	DoR (.47)*	DoR (.46)*
Neuroticism	DoR (.40)	QoI (.37) DoR (.45)* DoL (.44)

that extraverts' attention is directed outward, away from themselves [11]. These results for extraversion show how social psychology research is translated to the context of human-agent interaction. With regard to agreeableness, we found that more agreeable subjects reported higher DoR to both agents, which is also supported by [7], and to existing work in human-agent interaction and perception of rapport [12], which is not surprising, since agreeableness is associated with friendliness, warmth, and sociability. People with higher degree of neuroticism reported higher DoR, QoI and DoL only to agent Poppy. People high in conscientiousness reported higher QoI only for sad Obadiah. With regard to openness to experience, no significant results can be reported.

We would also like to drive comparison of our results to existing work on the influence of extraversion trait on the perception of virtual characters with respect to similarity rule, and opposite, complementary rule introduced in Section 1. Our results show two correlation effects of DoL and Big 5 traits for both agents: Obadiah, who is shown to have high neuroticism, and Poppy, who is shown to have high extraversion [17]. Extraverted people in our study show tendency to like Obadiah (for Poppy no relation is found), whereas more people with high neuroticism show tendency to like Poppy (for Obadiah no relation is found). These results show support for the complementary likeness rule.

4.2 Nonverbal Cues and Interaction Experience

We study the individual association between nonverbal features and measures of interaction experience, which are shown in Table 3. As a first result we found that interactions with agent Poppy results in higher cue utilization (18) than with agent Obadiah (10). One possible, albeit speculative, explanation for this could be that subjects freely expressed themselves in second interaction (with Poppy) because they knew what is expected from them. This has been confirmed for assessment of personality meaning that 'when strangers get to know each other, information contained early in the interaction may be less useful for making accurate personality assessments' (see [17], p. 315. for discussion).

Table 3. Significant Pearson correlation effects between interaction experience measures and nonverbal cues ($p < 0.05$, *$p < 0.01$), see cue acronyms in Section 3.2

	Obadiah	Poppy
QoI	ATD (-.35), ALSS (-.35) GES (-.36) # Cues: 3	TSL (.41), TST (.48)* F0 (m) (-.40), Val R0 (m) (.37), Loc R0 (m) (-.35), ALVS (-.36) HDO (-.35), BL (-.49)* # Cues: 8
DoR	ATD (-.36) ALSS (-.49)* BL (-.42) # Cues: 3	TST (.36), Val R0 (m) (.37), BL (-.47)*, SR (-.35), ST (-.46)* HDO (-.49)*, HDS (.38) # Cues: 7
DoL	ATD (-.35) ALSS (-.51)*, FTS (-.41) FSO (-.41) # Cues: 4	Loc R0 (std) (-.35) ALVS (-.36) HDO (-.38) # Cues: 3

Another possible explanation of this phenomena could lie in design of our study: self-reported interaction measures in second interaction (with Poppy) could be affected by subject's experience and measures reported after interaction with Obadiah. The issues and proposed solutions are discussed in Conclusions section.

With regards to QoI, we expected that audio cues will be more significant than visual cues we extracted, as we do not take into account any facial expressions (e.g. confuse or surprise). This was the case, more significant for agent Poppy, with results showing that longer speaking time and more turns mean higher QoI, which is not surprising, taking into account that SALs are designed to induce interaction. Besides, lower pitch, higher value of autocorrelation peaks (louder speech), lower location of autocorrelation peaks, and lower average length of voiced segment also show higher QoI for Poppy. Among visual cues, back leans and head oriented away were found to be significant. Results show how people who lean back more and 'look away from screen' more are not having high QoI with Poppy. Back leans in this case may indicate boredom or lack of interest, whereas gaze in psychological literature serves both as regulator of conversation flow and indicator of attitude. Several studies have confirmed how frequent amount and length of gazing communicates positive attitude towards conversational partner (see [15] for an overview), so in our case, more frequent head directed outwards may also indicate lack of interest. lthough in our case only head direction is computed, our results show that it is an acceptable approximation to eye contact in this scenario. For Obadiah QoI, we also got three significant results, showing that people who make less gestures and whose speech segments and turns last shorter are reporting higher QoI. Shorter speech segments and average turn duration, which are also reported for higher DoL and DoR for Obadiah, indicate that the interaction is indeed two ways, the agent responses to the subject, which could explain the high QoI. After the experiment, some subjects reported how 'they wanted to cheer up Obadiah', so these features could also

indicate how subjects who reported higher QoI, DoR and DoL felt an empathy with sad Obadiah. To support this theory, linguistic content of the speech could be analyzed.

For Obadiah, people who also lean back a lot show lower DoR. In case of Poppy, people who take more turns, lean back less, lean sideways right less, speak louder, touch themselves less, look away from screen less show higher DoR. These can all be identified as signals of interest. These features are also found to be significant in human-human interaction ([7,15,18]). Though, in case of sideways leans Mehrabian argues how reclining angle is important to differentiate positive and negative attitude.

With regards to DoL, in the case of Obadiah, only vocal features of the subjects were found to be significant. Average turn duration, average length of speaking segments, lower fraction of speaking time and fraction speaking over are related to higher DoL, which means people who speak less tend to like Obadiah more. With regards to visual nonverbal cues and DoL, we found only one significant result only for Poppy. People who like Poppy more, do not move their head away from screen a lot. This result is also related to findings on likeness in human-human interaction, showing how people show more direct eye contact to liked partner [24].

4.3 Regression Analysis

The task of predicting the interaction experience is addressed by building computational models for predicting the score of individual measures of perceived experience: QoI, DoR and DoL. For prediction task we used three different regression models: support vector regression (SVR), neural networks (NN) and kernel ridge regression (KRR). Each model is trained using double cross-validation (CV) approach in which for outer fold we used leave-one-out CV, and for inner fold we used 5-fold CV approach. The inner fold is used for parameter optimization. SVR and KRR models use RBF kernel. The models are trained with different feature sets: we experimented with (1) all extracted nonverbal behavior (NVB) cues, (2) all NVB cues and all personality traits (PT), (3) all PT, (4) significant NVB cues, (5) significant PT, and (6) significant NVB and PT (significant NVB cues and PT are shown in tables 2 and 3). Additionally, for all feature sets we also applied Principal Component Analysis (PCA), in order to reduce dimensionality of data.

Table 4 shows the results of our experiments, where we report the R^2 and Root Mean Square Error (RMSE). Among different feature sets and regression models that we have experimented with, we report the best results for each experience aspect. To stress the difference between experimented feature sets, we only show the results of the best regression model for a specific feature set.

One can first notice how for the best results are obtained when personality traits and nonverbal cues are combined, which boost the performance of each individual input source (PT and NVB used alone). QoI and DoR for both agents are predicted with R^2 of 0.3. Although the R^2 values found for QoI and DoR are on the low side, they are comparable to the results found in other studies

Table 4. Prediction results for Obadiah and Poppy with different feature sets (Personality traits (PT), nonverbal behavior (NVB), all vs. significant cues. For each feature set we only show results of the best regression model.

		Feature Set	Meth.	R2	RMSE
Obadiah	QoI	NVB+PT (sig.)	SVR	0.292	0.144
		PT (sig.)	SVR	0.158	0.157
		NVB+PT (all, PCA)	SVR	0.034	0.168
	DoR	NVB+PT (sig.)	SVR	0.34	0.137
		PT (sig.)	KRR	0.134	0.156
		NVB (sig.)	KRR	0.106	0.159
	DoL	NVB+PT (sig)	KRR	0.174	0.197
		NVB (sig.)	KRR	0.066	0.209
		PT (sig.)	SVR	0.016	0.215
Poppy	QoI	NVB+PT (sig.)	KRR	0.523	0.129
		NVB (sig.)	KRR	0.158	0.172
		PT (sig.)	SVR	0.015	0.186
	DoR	NVB+PT (sig.)	KRR	0.406	0.162
		NVB (sig)	KRR	0.158	0.192
		NVB+PT (all, PCA)	KRR	0.114	0.199
	DoL	NVB+PT (sig.)	SVR	0.322	0.200
		NVB+PT (all, PCA)	KRR	0.097	0.230

in social computing literature for predicting several other social aspects such as personality [1,3]. DoL is the weakest aspect of our prediction models: The highest R^2 for Obadiah is 0.174 and Poppy 0.322. With regards to the regression method, SVR and KRR have shown to produce similar results and they outperformed NN.

5 Conclusions

Our paper presented a study in which we attempt to analyze and predict the experience of interaction (or perceived interaction) with virtual characters. A novelty of our work is that we use a combination of nonverbal cues and personality traits to predict the experience. Best prediction results for all experience measures were obtained when nonverbal cues and personality traits are used together as features. The degree of rapport for agent Obadiah and quality of interaction for agent Poppy are the most predictable measures.

We examined self-reported personality traits and extracted nonverbal cues as descriptors of experience and found that personality traits are very significant features, as also reported in [28]. This is another confirmation how humans' personality shapes the experience of human-agent interaction, and how it should be assessed in virtual agents evaluation studies. Another finding related to our work shows how people with high agreeableness perceive strong rapport with an agent designed to build rapport [12]. Our results however suggest that characterization of an agent might not play a role in perceiving the rapport during interaction.

People who score high in agreeableness in our study reported higher rapport with both sad Obadiah and cheerful Poppy. This also points to prior findings on human-human interaction on how the presence of at least one agreeable member in the dyad results with higher rapport perceived with conversational partner [7]. We have also found that some of the extracted nonverbal features significant for describing experience are related to socio-psychological findings on affect and liking, such as body leaning and head orientation. Acoustic and paraliguistic features were shown as more meaningful descriptors than visual features, which is also phenomena observed for judgments of personality traits [8]. All nonverbal features, except hand activities are automatically exported. Cheap markerless motion capture system (MS Kinect v1.8) was found to be partially useful for the purpose of automatic social behavior analysis. Experiments with head information from the face tracking system were successful and we build head direction classifier. However, in subjective evaluation upper body tracking failed to produce reliable results for hand joints so hand activity was manually annotated. Moreover, additional information was required to build body-lean classifier from body tracking system.

A limitation of our study is experimental design. Limited by resources, we assigned our subjects to the same experimental conditions, in which they first completed the interaction with Obadiah, and then for Poppy. The questionnaire was applied after each interaction. One could expect that experience of the second interaction is affected by the first interaction. The same fact is the reason why we can not strongly support complementary likeness rule in HAI, which is suggested by our results. Besides, it has been observed how more significant nonverbal features, or higher cue utilization for interaction scores, were found in second interaction. As explained beforehand, subjects might have freely expressed themselves because they got accustomed to the SAL system.

To overcome somewhat arguable interaction scores we plan to crowdsource the annotation of observed experience from collected audio-visual data. Then, we will perform a study on comparison of self-reported and observed measures. Moreover, to improve prediction of experience additional nonverbal features are considered to be exported: to improve prediction of likeness e.g. eye shifts and eye contact could be useful [24], and to improve quality of interaction overlapped speech segments could be significant. Instead of using self-reported personality scores as inputs of our regression models, we plan to build computational models which predict personality from audio-visual data. For this task, a thorough study on prediction of personality from nonverbal cues from both human-agent interaction and self-presentation sequences will be done.

Acknowledgments. This work was partly funded by the Swiss National Science Foundation (SNSF) Ambizione project "Multimodal Computational Modeling of Nonverbal Social Behavior in Face to Face Interaction" (PZ00P2-136811) and by grants from the Croatian Science Foundation (CSF), and Pascal 2 Network of Excellence.

References

1. Aran, O., Gatica-Perez, D.: One of a Kind: Inferring Personality Impressions in Meetings. In: International Conference on Multimodal Interaction (ICMI), Sydney, pp. 11–18 (2013)
2. Argyle, M.: Bodily communication. Methuen (1988)
3. Biel, J.-I., Gatica-Perez, D.: The Youtube lens: Crowdsourced personality impression and audiovisual of vlogs. IEEE Transactions on Multimedia 15(1), 41–55 (2012)
4. Biel, J.-I., Aran, O., Gatica-Perez, D.: The Good, the Bad, and the Angry: Analyzing Crowdsourced Impressions of Vloggers. In: International Conference on Weblogs and Social Media, ICWSM (2012)
5. Buisine, S., Martin, J.C.: The influence of user's personality and gender on the processing of virtual agents multimodal behavior. Advances in Psychology Research, 1–14 (2010)
6. Cafaro, A., Vilhjálmsson, H.H., Bickmore, T., Heylen, D., Jóhannsdóttir, K.R., Valgardhsson, G.S.: First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 67–80. Springer, Heidelberg (2012)
7. Cuperman, R., Ickes, W.: Big five predictors of behavior and perceptions in initial dyadic interactions: personality similarity helps extraverts and introverts, but hurts disagreeables. Journal of Personality and Social Psychology 97(4), 667–684 (2009)
8. Ekman, P., Friesen, W., O'Sullivan, M., Cherer, K.: Relative importance of face, body, and speech in judgments of personality and affect. Journal of Personality and Social Psychology 38(2), 270–277 (1980)
9. Funder, D.C., Sneed, C.D.: Behavioral manifestations of personality: An ecological approach to judgmental accuracy. Journal of Personality and Social Psychology 64, 479–490 (1993)
10. Isbister, K., Nass, C.: 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)
11. Jung, C.: Psychological types. Harcourt, Brace, New York (1921)
12. Kang, S.-H., Gratch, J., Wang, N., Watt, J.H.: Agreeable people like agreeable virtual humans. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 253–261. Springer, Heidelberg (2008)
13. Knapp, M.L., Hall, J.A., Horgan, T.G.: Nonverbal Communication in Human Interaction, 8th edn. Cengage Learning (January 2013)
14. Lefter, I., Burghouts, G.J., Rothkrantz, L.J.M.: Automatic Audio-Visual Fusion for Aggression Detection Using Meta-information avss. In: 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, pp. 19–24 (2012)
15. Maxwell, G.M., Cook, M.W.: Postural congruence and judgements of liking and perceived similarity. New Zealand Journal of Psychology 15(1), 20–26 (1985)
16. McCrae, R.R., Costa, P.T.: A contemplated revision of the NEO Five-Factor Inventory. Personality and Individual Differences 36(3), 587–596 (2004)
17. McRorie, M., Sneddon, I., McKeown, G., Bevacqua, E., Sevin, E., Pelachaud, C.: Evaluation of four designed virtual agent personalities. Transactions on Affective Computing 3(3), 311–322 (2012)

18. Mehrabian, A.: Significance of posture and position in the communication of attitude and status relationships. Psychological Bulletin 71(5), 359–372 (1969)
19. Nass, C., Lee, K.M.: Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In: CHI 2000: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, pp. 329–336 (2000)
20. Nguyen, L., Marcos-Ramiro, A., Marron-Romera, M., Gatica-Perez, D.: Multimodal Analysis of Body Communication Cues in Employment Interviews. In: Proc. ACM Int. Conf. on Multimodal Interaction (ICMI), Sydney (December 2013)
21. Palmer, M.T., Simmons, K.B.: Communicating intentions through nonverbal behaviors conscious and nonconscious encoding of liking. Human Communication Research 22(1), 128–160 (1995)
22. Pentland, A.S.: Honest Signals: How They Shape Our World. The MIT Press (2008)
23. Pianesi, F., Zancanaro, M., Lepri, B., Cappelletti, A.: A multimodal annotated corpus of consensus decision making meetings. Language Resources and Evaluation 41(3), 409–429 (2007)
24. Ray, G.B., Floyd, K.: Nonverbal expressions of liking and disliking in initial interaction: Encoding and decoding perspectives. Southern Communication Journal 71(1), 45–65 (2006)
25. Sanchez-Cortes, D., Aran, O., Jayagopi, D., Schmid Mast, M., Gatica-Perez, D.: Emergent Leaders through Looking and Speaking: From Audio-Visual Data to Multimodal Recognition. Journal on Multimodal User Interfaces Special Issue on Multimodal Corpora 7(1-2), 39–53 (2013) (published online August 2012)
26. Sun, X., Nijholt, A., Truong, K.P., Pantic, M.: Automatic understanding of affective and social signals by multimodal mimicry recognition. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part II. LNCS, vol. 6975, pp. 289–296. Springer, Heidelberg (2011)
27. Tapus, A., Mataric, M.J.: Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In: Emotion, Personality, and Social Behavior, pp. 133–140. AAAI (2008)
28. von der Pütten, A.M., Krämer, N.C., Gratch, J.: How our personality shapes our interactions with virtual characters - implications for research and development. In: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 208–221. Springer, Heidelberg (2010)
29. Vacharkulksemsuk, T.B., Fredrickson, L.: Strangers in sync: Achieving embodied rapport through shared movements. Journal of Experimental Social Psychology 48(1), 399–402 (2012)
30. Vijayasenan, D., Valente, F., Bourlard, H.: An Information Theoretic Combination of MFCC and TDOA Features for Speaker Diarization. IEEE Transactions on Audio Speech and Language Processing 19(2) (2011)
31. Waldron, J.: Judgement of like-dislike from facial expression and body posture. Perceptual and Motor Skills 41(3), 799–804 (1975)