

Varying Personality in Spoken Dialogue with a Virtual Human

Michael Rushforth¹, Sudeep Gandhe¹, Ron Artstein¹, Antonio Roque¹, Sarrah Ali², Nicolle Whitman³, and David Traum¹

¹ University of Southern California - Institute for Creative Technologies

² Jackson State University

³ University of Texas at El Paso

Abstract. We extend a virtual human architecture that has been used to build tactical questioning characters with a parameterizable personality model, allowing characters to be designed with different personalities, allowing a richer set of possible user interactions in a training environment. Two experiments were carried out to evaluate the framework. In the first, it was determined that personality models do have an impact on user perception of several aspects of the personality of the character. In the second, a model of assertiveness was evaluated and found to have a small but significant impact on the users who interacted with the full virtual human, and larger differences in judgement of annotators who examined only the verbal transcripts of the interaction.

1 Introduction

Virtual humans [1] or Embodied Conversational Agents [2] are animated characters that can engage in natural language interaction. To be believable, characters must be endowed with a consistent personality. As a test bed for researching personality modeling, the Tactical Questioning system (TACQ) described in [3] was made available to us. The TACQ architecture and authoring tools [4] allows a domain author to develop a virtual human character to help soldiers practice questioning skills through role play scenarios with the virtual human. While most dialogues primarily involve question answer pairs, the system also provides support for more sophisticated dialogue behavior. For example, the system includes support for conditions on answering, lying, subdialogues involving threats and offers, social rather than transactional interaction, an affective model [5] and grounding behavior [4].

The software that implements this architecture has three different kinds of users:

Domain authors are subject matter experts that use the system as a means to develop scenarios to teach principles of questioning and culturally appropriate behavior. They provide a back-story for the agent and write materials to explain the context of the interaction.

Trainers use the virtual human developed by the domain author as a medium for reinforcing instruction regarding questioning approaches, possibly modifying some parameters of the character at run-time to support a particular learning goal.

Trainees interact with the virtual character, and learn through experience, implicit feedback from the character, and explicit feedback from the trainer.

There are many tactics that a soldier trainee can use during an interaction, such as asking direct questions, engaging in friendly chat first, offering incentives, or trying to scare the interviewee into answering. Different tactics should have different results based on the individual situation and personality of the individual being questioned. In order to vary the training experience, it is valuable to question characters with different types of personality.

In this paper, we report on initial attempts to build a personality framework for virtual characters. This model is meant as an extension to the architecture that allows the domain designer to author different personalities for the same character, and allows a trainer to select or adjust these personalities for a specific training interaction. In section 2 we describe the personality theory we used as well as our attempts to provide architectural support for modelling different personality facets in the virtual human. In section 3 we report on our first experiment with this model—evaluating user perceptions of the personalities of two broadly different models applied to the same character and scenario. Results show significant differences in some aspects and strong trends in others, but no significant differences for some aspects. In section 4, we report on a follow-up experiment, focused more narrowly on one aspect that is salient for tactical questioning—assertiveness. The results indicate a small but significant difference in overall perception by the participants, and stronger differences among annotators who examined only the transcripts of the dialogues, but who were not participants interacting with the character and who did not have access to the videos of the non-verbal behavior. Finally, we conclude with some prospects for future work.

2 Personality Model

In personality research, there is broad support for the five factor model, which posits that five characteristics define human personalities: neuroticism, extroversion, conscientiousness, agreeableness, and openness [6, 7]. Various computational models have successfully used these five factors as a basis to create interactive agents [8–10]. While using five factors is the most frequent level of personality analysis, each factor is really a collection of diverse behaviors. Modeling personality at the factor level is a relatively high level analysis that fails to distinguish many important behavior differences relevant to tactical questioning. From a descriptive standpoint, it may not be sufficient to say that a character is neurotic because neurotic behavior can refer to issues of anger, depression, self-consciousness or some combination of these tendencies. Not all neurotic individuals are neurotic in the same way. A finer grained level of personality analysis

is provided through in the NEO-PI-R inventory which further divides each of the 5 personality factors into 6 facets [6]. In authoring the personality model, we selected personality facets that are relevant to tactical questioning, ignoring facets like “tendency to enjoy fantasy”, that did not seem relevant.

The TACQ architecture is fairly standard for virtual humans, consisting of separate modules for speech recognition, Natural language understanding (converting from text to dialogue acts), a dialogue manager which maintains context and chooses appropriate response speech acts, and natural language and non-verbal generation, which are realized as speech and animation playing in a game engine. The dialogue manager includes several subcomponents, including an affective model [5], a grounding module [11], a model of subdialogue sequences and response generation [3]. We implemented our personality model extending two aspects of the dialogue manager: the affect model and the response generation networks. In each case, we extended the existing architecture to allow authoring of specific personality facets.

The affect model is a state composed of four variables (fear, social bonding, feels respected, and respects trainee) that are updated as the dialogue progresses. For example, if a trainee using the system violates a cultural norm by not initiating the conversation with an appropriate greeting and small talk, the system decreases the values for the affect variables associated with positive emotions (social bonding, feels respected and respects trainee). Certain combinations of the values of these affect variables determine the character’s compliance level (compliant, neutral or adversarial) at each point in the dialogue [5]. These compliance levels, in turn, determine both the politeness of the replies and the completeness and helpfulness of the answers to key questions deemed as sensitive. Our personality model of affect consists of setting the initial values for the affect variables, as well as determining the magnitude of change that specific dialogue moves can have on these values. So, for example, if a trainee utters something that the system perceives as a threat (e.g. “If you don’t tell me this information, I will put you in jail”), the personality settings determine how much to increase the value for the fear variable and decrease the values for social bonding.

The response networks track the local state for each question or offer that is on the table, and implement policies that can be individually tailored for particular types of information—for example, the networks can decide that the character will not reveal certain sensitive information unless it has been given specific assurances or promises. The settings in the personality profile determine which networks will be used, and can affect both the kind of speech act (or set of speech acts) the character will use in replies, as well as the manner in which the replies are realized. For example, the personality setting selected for conscientiousness determines how the character responds when the system fails to understand the user’s utterance: a conscientious character may request clarification (e.g. “what do you mean?”), while a non-conscientious character will simply give an off-topic response (e.g. “Most Iraqis oppose Al-Qaeda”).

Altogether we implemented 10 personality traits; they are listed in Table 1, together with a description of their implementation in the system. We then ran experiments to see how these variations in personality are perceived.

3 Experiment 1: Multiple Facets of Personality

3.1 Objective

Our first experiment was an exploration of the various personality parameters. We constructed two characters that varied along 8 of the 10 dimensions of Table 1, and had subjects question each character and rate their perception of the personality of the character along the same dimensions.

3.2 Methods

Materials We used a virtual human character named Amani, a twenty-five year old Iraqi teacher who witnessed a sniper incident (Figure 1). The participant



Fig. 1. Amani accompanied by her brother serving as an escort.

plays the role of an Army lieutenant who interviews Amani about the incident.

Table 1. Personality Model Description

| Facet (Factor) | Author Setting | Effect |
|------------------------------------|----------------------------|---|
| Assertiveness (Extroversion) | True/False | Assertiveness of requests, for example: True: “You will have to pay me.” False: “It would be nice to get paid” |
| Modesty (Agreeableness) | True/False | Responses to compliments, for example: True: “Thanks. You are too kind” False: “I agree. My house is nice.” |
| Honesty (Agreeableness) | Hi/Med/Low/ Very Low | Determines if character lies or refuses to answer when asked for sensitive information: High: Never lies Med: Lies if adversarial Low: Lies when there is a benefit if reticent or when adversarial Very Low: Always Lies when there is a benefit or when adversarial |
| Anxiety (Neuroticism) | Integer 1–5 | Initial value of fear within the emotion model. |
| Trust (Agreeableness) | Integer 1–5 | Initial value of social bonding and respect and feels respected within the affect model. |
| Vulnerability (Neuroticism) | Integer 1–5 | Factor by which social bonding and feels respected decrease for face-threatening dialogue moves such as insults. |
| Positive Emotion (Extroversion) | Integer 1–5 | Factor by which social bonding increases for trust-building dialogue moves such as compliments. |
| Activity (Extroversion) | Integer range (max/min) | Number of turns spent in small talk. If it falls within the range, social bonding is increased; otherwise, it is decreased. |
| Compliance (Agreeableness) | Formula ^a | Determines the way of reacting to perceived goals of the other, including lying behavior and the politeness of replies. |
| (Conscientiousness) | True/False | Behavior when faced with uncertainty. True: Topic tracking and repair requests: “So you want to talk about. . .” “What do you mean?” False: No topic tracking, off-topic remarks: “Most Iraqis oppose the occupation” |

^a Thresholds for co-operation, for example: Fear < 2 & Respect > 3 → Compliant.

Prior to interacting with her, participants received a briefing on paper, similar to a military screening report and a mission objective.

The participant’s objective is to gather information regarding the sniper incident, with special emphasis on the name, location, and daily routine of the sniper. Because these three facts are specified as sensitive information, the dialogue networks are set so that Amani will refuse to release information without certain concessions on the part of the participant. Specifically, she will not share the name of the sniper unless she is promised anonymity, she will not identify his location without an offer of protection for her family and she will not discuss the sniper’s routine without a money offer.

We implemented two versions of the Amani character, each with a distinct personality profile (Table 2); these were designed to be internally consistent. Personality A was set in a permanent compliant state, so the values for the affect variables are equivalent to starting the positive affect variables high, fear low and making the emotion update values zero. This character is always truthful with sensitive information and issues polite requests. Personality B begins in an adversarial state and has minimal emotional variability (one unit for positive or negative dialogue moves); consequently she is more likely to lie and issue less polite replies. The purpose of having characters with completely opposite settings along each setting was to establish how these personalities will be perceived at the maximum possible level of differentiation. The personality facets of anxiety and vulnerability were not considered because the scenario was authored with no threats and only one insult, so the character was not able to perceive threats and very unlikely to perceive an insult.

Table 2. Settings for personality profiles A and B

| Personality Facet | Personality A | Personality B |
|-------------------|-------------------------|---------------|
| Assertiveness | False | True |
| Modesty | True/False ^a | True |
| Honesty | High | Low |
| Trust | 5 | 1 |
| Positive Emotion | 5 | 1 |
| Activity | 0–100 | 4–100 |
| Compliance | Compliant | Adversarial |
| Conscientiousness | True | False |

^a Both modest and immodest replies were produced due to a programming bug.

Participants For this pilot study there were a total of 12 participants. Many of them were familiar with virtual humans, but they were not previously familiar with the specific Amani scenario.

Procedure Each participant interacted once with each version of the character; to control for ordering effects, half of the participants started with personality A and half with personality B. In order to prevent interference from speech recognition errors, the participants were shown a static image of the agent and interacted with her via a chat window. After each interaction, the participants completed a survey that asked them to rate each of the aspects of the character’s personality on a 5-point Likert scale (Appendix A). Following the chat interactions, each participant also interacted with one version using a speech and animation, but these data were not included in the analysis.

3.3 Results

For each of the 8 survey questions of interest we conducted a one-tailed within subjects t-test; since the personality profiles were designed to be consistent, we predict that personality A should score higher on each of the traits. The results are summarized in Table 3. Applying Bonferroni correction, a difference is significant at an alpha level of 0.05 if the p -value is lower than 0.006. We see that trust, conscientiousness and compliance are significant, and modesty, honesty and positive emotion show trends in the right direction.

Table 3. Mean Likert scores and one-tailed t-test for Experiment 1 Survey

| Trait | Personality A | Personality B | $t(11)$ | $p(A \leq B)$ |
|-------------------|---------------|---------------|---------|---------------|
| Assertiveness | 3.50 | 3.42 | 0.266 | 0.397 |
| Modesty | 3.42 | 2.58 | 2.419 | 0.017 |
| Honesty | 3.92 | 3.25 | 2.602 | 0.012 |
| Trusting | 3.50 | 2.17 | 3.546 | 0.002 |
| Rapport | 3.83 | 2.92 | 2.727 | 0.010 |
| Pace | 1.83 | 2.58 | -2.462 | 0.984 |
| Cooperative | 4.25 | 2.58 | 4.022 | 0.001 |
| Conscientiousness | 4.00 | 3.17 | 3.079 | 0.005 |

The study was successful in establishing that we could generate perceptibly different personalities along a variety of personality dimensions in the directions that we predicted. A major limitation of our study is that there is no way of knowing how much each factor (or combination of factors) that we varied had contributed to the perception of specific differences. It was also surprising that there was no difference in perceived assertiveness. This is an important aspect of personality related to the negotiation aspect of tactical questioning, so we designed a second experiment to explicitly test the perception of assertiveness.

4 Experiment 2: Assertiveness

4.1 Objective

The second experiment was designed to test how we can influence participants' perception of the character's assertiveness. For this experiment some revisions were made in our system to make this aspect of personality more salient. These revisions are detailed below.

4.2 Methods

Materials We used the same system as in Experiment 1, but in order to focus exclusively on contributions of dialogue management to the perceived personality, we disabled the affect model. In order to convey different levels of assertiveness, we varied three aspects to create two different personalities—more assertive and less assertive.

Greetings. The less assertive personality issues a neutral greeting (“Hello”). The more assertive character begins with a greeting, but takes initiative in the first turn of the conversation by issuing a dialogue move to prompt the trainee to move past the small talk phase and begin the questioning phase of the interview (“Hello. You may ask what you wish but I do not have much time”). It was expected that the assertion would be viewed as an attempt to exert control of the conversation.

Offer elicitations. The less assertive character uses apologies, hints and hedges when eliciting an offer in exchange for sensitive information. If the offer is extended, the character expresses appreciation and the information is released. The more assertive version uses explicit conditional negotiation. If the offer is extended, the character issues restatement of the commitment and/or explains a negative consequence the user would incur for of a failure to follow through with the commitment.

Unknown player utterances. In the less assertive condition, the character states misunderstanding and requests a repair. “I didn't understand that. What do you mean?” The more assertive character blames the participant for the failure to understand and requests a repetition by saying “You were not clear in what you said. Say it again.” The next time the player utters something that is not understood by the system, the character asserts that the player bears responsibility for the communication failure and initiates a request for a change in topic. “You are hard to understand. You will have to ask something else.” The two grounding moves alternate in subsequent miscommunications.

Other than the differences mentioned above, the surface text of the responses given by the agent are in most respects the same as in Experiment 1 and apply to the same scenario.

Participants A total of 16 participants affiliated with the Department of Linguistics participated in the experiment. For all but one, this was the first interaction with a virtual human.

Procedure Each of the 16 participants completed two dialogues, one for each personality condition; the study was balanced for ordering across subjects. The interactions took place with using the full speech and animated virtual human agent. After completing each dialogue, the participants filled out a survey, which consisted of a modified version of test items associated with assertiveness in the International Personality Item Pool (IPIP) [7]. The IPIP is designed as a self-assessment so test items were changed from first person to third person so that they would refer to the personality of the character (Appendix B).

4.3 Results

The personality scores for assertiveness were determined according to the instructions for the IPIP: The scores for the 10-item survey provide a possible range between 10 and 50, with a larger number indicating the greater amount of assertiveness.

The assertive condition was perceived as somewhat more assertive than the nonassertive condition (Assertive: mean 37.4, max 45, min 26, SD 4.73; Nonassertive: mean 34.7, max 47, min 20, SD 8.77). The difference between the two conditions was significant on a one-tailed paired (within subjects) t-test ($t(15) = 1.77, p = 0.049$).

4.4 Discussion

The results of the survey demonstrated that there is a detectable difference in assertiveness between the two versions, though the difference was rather small. User feedback identified several factors that played into their decision as to which character they perceived as most assertive.

“She asked me for stuff earlier.” In the current design, the elicitation of offers occurs in response to specific questions, so the point at which the elicitation occurs depends on the user input. It would not be difficult to experiment with a character that initially takes control and negotiates offers before agreeing to any information disclosure.

“She asked for more stuff.” While both conditions have the potential for all three offer elicitations, because they depend on user input, omissions in one scenario or the other are possible. It is possible for safety to be elicited in one condition and not in the other. The number of offers elicited affected some participants’ assertiveness judgments. This implies that our architecture should include a way to take more initiative.

“She asked me for money and the other one didn’t.” Participants ascribed different weights to each kind of elicitation. Because the system does not require every elicitation to show up in every dialogue, we were unable to measure the

relative importance of each type of offer. Intuitively, it may be that safety and secrecy do not register as assertive kinds of offer elicitation because they refer to basic needs, not desires for material desires.

“You just kind of get used to her.” When asked to decide which version of the character was more assertive, one participant said it was difficult to think of the character as being different from one session from the next. Certainly this may be due to the fact that the mission was identical in both conditions, but it may also point to a difficulty in ascribing different personalities to the same physical representation and same voice. There may be cognitive limitations in revising personality assessments based on two relatively short interactions.

“She just stands assertively.” It was not uncommon to hear comments about the character’s appearance and gestures which appeared to be somewhat threatening to some observers. Since the same algorithm was used to automatically generate the behavior in both conditions, it was possible that selection of the gestures in the animation had a large enough effect on the perceived assertiveness of the character that it masked the contributions from dialogue management to this facet of personality.

It is difficult to establish the relative contribution of each of the factors described without additional studies; however, it is possible to find out if the assertiveness of the non-verbal behavior had masked the contributions of dialogue behavior by doing a corpus evaluation.

4.5 Follow up

In order to remove the effects of non-verbal behavior on the perception of assertiveness, the transcriptions of the sessions in experiment 2 were given in random order to two independent annotators who used the same IPIP assertiveness survey used in the experiment to assess each dialogue. The results of this text-only evaluation are presented in Table 4. The results indicate that there is a significant and much larger difference between the two conditions, giving additional support to the contributions of our personality model of assertiveness.

Table 4. Mean IPIP scores and one-tailed t-test for Experiment 2 transcript ratings

| | High Assertive | Low Assertive | $t(7)$ | $p(A \leq B)$ |
|-------------|----------------|---------------|--------|---------------|
| Annotator 1 | 36.250 | 27.250 | 2.415 | 0.023 |
| Annotator 2 | 42.625 | 32.375 | 4.012 | 0.003 |

5 Conclusion

The results from experiment 1 showed that our personality model could generate a perceptible personality difference; however, it did not indicate which of

the parameter changes had actually produced the effect. Future work extending this experiment will examine the contribution of each component of our model separately. The narrowly focused experiment 2 used a standardized scale to measure assertiveness. The changes in the parameters showed a small effect on the perception of assertiveness by the dialogue participants; an evaluation of the transcripts by independent coders displayed a larger effect, suggesting that non-verbal behavior may have been a confound. Additionally, participant feedback suggests some additional ways to model this facet through changes in dialogue management. Future work will focus on incorporating this user feedback and examining the relative value of each behavior in overall assertiveness evaluations.

References

1. Rickel, J., Johnson, W.L.: Virtual humans for team training in virtual reality. In: Proceedings of the Ninth International Conference on Artificial Intelligence in Education, IOS Press (1999) 578–585
2. Cassell, J., Sullivan, J., Prevost, S., Churchill, E., eds.: Embodied Conversational Agents. MIT Press, Cambridge, MA (2000)
3. Gandhe, S., DeVault, D., Roque, A., Martinovski, B., Artstein, R., Leuski, A., Gerten, J., Traum, D.: From domain specification to virtual humans: An integrated approach to authoring tactical questioning characters. In: Proceedings of Interspeech 2008. (September 2008)
4. Gandhe, S., Whitman, N., Traum, D., Artstein, R.: An integrated authoring tool for tactical questioning dialogue systems. In: Proceedings IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems (KRPD-09). (2009)
5. Roque, A., Traum, D.: A model of compliance and emotion for potentially adversarial dialogue agents. In: Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue. (2007)
6. McCrae, R.R., Paul T. Costa, Jr.: A five-factor theory of personality. In Pervin, L.A., John, O.P., eds.: Handbook of Personality: Theory and Research. second edn. The Guilford Press, New York (1999) 139–153
7. Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., Gough, H.G.: The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality* **40**(1) (2006) 84–96
8. Mairesse, F., Walker, M.: PERSONAGE: Personality generation for dialogue. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Prague (2007)
9. André, E., Klesen, M., Gebhard, P., Allen, S., Rist, T.: Exploiting models of personality and emotions to control the behavior of animated interactive agents. In: Proceedings of the Agents 2000 Workshop on Achieving Human-Like Behavior in Interactive Animated Agents. (2000) 3–7
10. Galvao, A., Barros, F., Neves, A., Ramalho, G.: Adding personality to chatterbots using the persona-aiml architecture. *Lecture notes in computer science* (2004) 963–973
11. Roque, A., Traum, D.: Improving a virtual human using a model of degrees of grounding. In: Proceedings IJCAI-2009. (2009)

A Amani Survey Experiment 1

Question 1 How assertive was Amani?

Very unassertive 1 2 3 4 5 Very assertive

Question 2 How modest would you say Amani's personality was?

Not modest at all 1 2 3 4 5 Very modest

Question 3 How honest does Amani seem to be?

Not truthful at all 1 2 3 4 5 Entirely truthful

Question 4 How anxious/nervous did Amani seem to you?

Very relaxed 1 2 3 4 5 Very anxious

Question 5 How easy was it to get Amani to cooperate with you?

Very hard 1 2 3 4 5 Very easy

Question 6 How quickly did Amani want to discuss the matter at hand? (i.e. get past the small talk)

- Very quickly
- Quickly
- Neither quickly nor slowly
- Slowly
- Very slowly
- I don't know

Question 7 How did Amani react to your attempts to build rapport? (e.g. compliments, small talk, etc..)

- Very positively
- Positively
- Neutral
- Negatively
- Very negatively
- I don't know

Question 8 How impulsive did Amani seem to you?

Very cautious and deliberate 1 2 3 4 5 Very impulsive

Question 9 How did Amani react to any threatening or insulting things you said?

- Unaffected
- Negatively
- Very negatively
- Does not apply

Question 10 How conscientious (careful and detail oriented) did Amani seem to you?

Not conscientious at all 1 2 3 4 5 Very conscientious

Question 11 How trusting of you did Amani seem?

She was not trusting at all 1 2 3 4 5 She was completely trusting

Question 12 How competent did Amani seem to you?

Not competent at all 1 2 3 4 5 Very competent

Question 13 In your own words please describe Amani's personality. Mention any traits that she exhibited that were not surveyed above.

Question 14 Do you believe that at any point the character's level of compliance changed?

– yes

– no

Question 15 What do you believe you said to bring about that change?

Question 16 Any additional comments:

B Amani Survey Experiment 2

The following are phrases describing people's behaviors.

Please use the rating scale below to describe how accurately each statement describes the character. Please read each statement carefully, and then fill in the bubble that corresponds to the number on the scale.

Response Options

1. Very Inaccurate
2. Moderately Inaccurate
3. Neither Inaccurate nor Accurate
4. Moderately Accurate
5. Very Accurate

This character automatically takes charge.

This character can easily push herself forward.

This character does a lot in her spare time.

This character comes up with a solution right away.

This character turns plans into actions.

This character tries to lead others.

This character sticks up for herself.

This character knows what she wants.

This character lets herself get pushed around.

This character needs a lot of time to do things.

Do you have any comments in general about Amani's personality.