Impact of Personality on Nonverbal Behavior Generation

Ryo Ishii
rishii@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, PA

Chaitanya Ahuja
cahuja@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, PA

Yukiko I. Nakano
y.nakano@st.seikei.ac.jp
Carnegie Mellon University
Pittsburgh, PA

Louis-Philippe Morency
morency@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, PA

ABSTRACT

To realize natural-looking virtual agents, one key technical challenge is to automatically generate nonverbal behaviors from spoken language. Since nonverbal behavior varies depending on personality, it is important to generate these nonverbal behaviors to match the expected personality of a virtual agent. In this work, we study how personality traits relate to the process of generating individual nonverbal behaviors from the whole body, including the head, eye gaze, arms, and posture. To study this, we first created a dialogue corpus including transcripts, a broad range of labelled nonverbal behaviors, and the Big Five personality scores of participants in dyad interactions. We constructed models that can predict each nonverbal behavior label given as an input language representation from the participants' spoken sentences. Our experimental results show that personality can help improve the prediction of nonverbal behaviors.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

text-to-body motion, big five, personality, nonverbal behavior generation

ACM Reference Format:
Ryo Ishii, Chaitanya Ahuja, Yukiko I. Nakano, and Louis-Philippe Morency. 2020. Impact of Personality on Nonverbal Behavior Generation. In *IVA '20: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20), October 19–23, 2020, Virtual Event, Scotland Uk*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3383652.3423908

1 INTRODUCTION

In human communication, physical nonverbal behaviors, such as nodding, head posture, facial expressions, hand gestures, and upper-body posture, are used to express emotions and intentions [37, 39]. Therefore, it has been shown that enabling a virtual agent or robot to express appropriate nonverbal behaviors not only improves their natural appearance but also promotes conversation. For example, nonverbal behaviors accompanying an utterance have the effect of strengthening the persuasive power of that utterance, making it easier for the other party to understand the content of the utterance [27]. Thus, many studies have been conducted to enable virtual agents to automatically generate appropriate nonverbal behaviors [2, 4, 5, 7, 8, 13, 16, 20, 22, 23, 25, 31, 33, 39, 41].

Human nonverbal behaviors vary greatly depending on the personality of the individual [29, 35]. Some studies demonstrate that the Big Five affects some nonverbal behaviors [10, 18, 24, 35, 38]. Other previous studies demonstrated that humans can perceive differences in the Big Five personality traits of agents on the basis of differences in their nonverbal behaviors [32, 38, 40]. This supports the argument that humans would have a more natural and positive impression on an agent if it generates nonverbal behaviors consistent with its personality.

In this paper, we study how personality traits impact the generation of individual nonverbal behaviors. Our goal is to create a new generation model that can predict nonverbal behaviors that reflect an agent's personality taken as input. To make this study possible, we create a Japanese-dialogue corpus including transcripts, labelled nonverbal behaviors, and Big Five scores of participants. This corpus can be used to construct models that can predict nonverbal behaviors on the basis of compiled representations of input spoken languages. We study the impact of adding Big Five personality information to such prediction models. We especially focus on four types of nonverbal behavior: eye gaze, head nod, hand gesture, and upper-body posture.

2 RELATED WORK

2.1 Nonverbal Behavior Generation

Many attempts have been made to generate skeleton poses of humans during speaking, especially using speech information [2, 4, 5, 13, 14, 16, 23, 25, 39, 41]. These techniques can directly generate human-like nonverbal behaviors of agents designed with skeletal information similar to humans. Moreover, several studies have estimated nonverbal behaviors using spoken languages

[7, 8, 20, 22, 31, 33]. These studies worked on estimating the labels of nonverbal behaviors. The technique of using spoken languages is expected to be widely used for virtual agents that can talk with users using voice or text modalities [19]. Also, appearances and controllable body joints vary greatly depending on the design of the virtual agents. Therefore, a model that can output nonverbal behavior labels, which indicate what kind of behavior should be performed, could be used for a wide variety of virtual agents.

Our work focuses on how Big Five personality information can help models in automatically predicting nonverbal behavior labels using spoken language for the first time.

2.2 Relationships between Personality Trait and Nonverbal Behavior

Some research has demonstrated a relationship between the Big Five personality traits, which are among the most well-accepted indicators of personality and the most commonly used model of personality in academic psychology, and nonverbal behavior with statistical analysis. In detail, spatial nonverbal attributes such as body attitude, gesture amplitude or expansiveness, behavior direction, smoothness, and fluency have been shown to be key indicators of personality [29, 35]. Some research reported that extraversion affects some nonverbal behaviors. For example, leaning the upper body forward, tilting the head, energetic physical movement, and eye contact and gesture rates are also positively correlated with extraversion [10, 18, 24, 29, 35]. Moreover, Smith and Wang et al. [38, 40] demonstrated that differences in hand and arm movements depended on all Big Five indicators. These previous research results support the validity of our approach of using the Big Five for automatically generating individual nonverbal behaviors.

2.3 Personality Trait and Nonverbal Behavior in Virtual Agent

Some research [32, 38, 40] experimentally implemented different nonverbal behaviors in virtual agents and evaluated how humans can perceive personality from nonverbal behaviors. They demonstrated that a human can perceive differences in personality traits (Big Five) on the basis of the differences in the nonverbal behavior of a conversational agent. These results support the belief that humans have a more natural and positive impression toward an agent when appropriate nonverbal behaviors are generated that are related to the personality expressed by the agent. Therefore, for nonverbal behavior generation for virtual agents, it is important to consider personality traits.

3 CORPUS

To our knowledge, there is no corpus data that has time-series data of participants' verbal and various nonverbal behaviors and the Big Five for Japanese dialogue. Therefore, we first constructed a new dialogue corpus that includes these.

3.1 Verbal and Nonverbal Behaviors

We collected a corpus that includes verbal and nonverbal behaviors and personality trait (Big Five) information from human dialogue. We recorded 28 face-to-face conversations with 11 pairs of people. Each pair had 2 or 3 conversation sessions. The participants were Japanese males and females in their 20s to 50s who had never met before. They sat facing each other.

Table 1: List of labels for predicted nonverbal behaviors

Behavior types	#	Label list
Eye gaze	5	person head, person body, center side, under side, upper
Head nod	6	none, 1, 2, 3, 4, more than 5
Head direction (yaw)	3	center, side_s, side_l
Head direction (roll)	3	center, tilt_s, tilt_l
Head direction (pitch)	4	center, under_s, under_l, upper
Hand gesture	5	none, iconic, metaphoric, beat, others
Upper-body posture	7	center, forward_ss, forward_s, forward_m, forward_l, backward_s, backward_l

To ensure that there was a variety of conversation topics, we adopted two kinds of conversation. One was an explanation task with a famous cartoon called "Tom & Jerry." Just before starting a conversation, the participants watched a cartoon short story a few minutes long, and the characters did not speak. In each conversation session, one participant explained the content of the cartoon to the conversational partner within ten minutes. At any time during this period, the partner could freely ask questions about the content. The second conversation topic was a discussion on general topics such as tax and social welfare balance. The participants had no restrictions on remarks or topic changes.

We recorded the participants' voices with a lightweight headset microphone and video recorded the entire conversation. We also took a video showing the upper body of each participant (recorded at 30 Hz). In each conversation session, data on the utterances and major nonverbal behaviors were extracted during the ten-minute period (280 minutes in total) as follows (a list of extracted nonverbal behaviors is shown in Table 1).

Utterances. A professional annotated utterance units using the inter-pausal unit (IPU). Utterance intervals were manually extracted from speech waves. A portion of an utterance followed by 200 ms of silence was used as the unit of one utterance. We collected 6967 IPUs. In Japanese, a "bunsetsu," which consists of at least one content word and zero or more function words, is the minimum unit that conveys semantic content. In this study, we will refer to "bunsetsu" as "phrase" hereafter, but it should be noted that "bunsetsu" in Japanese is a smaller unit than a "phrase" in English. We used J-tag [11], which is a general morphological analysis tool for Japanese, to divide an IPU into phrases. We collected a total of 19,847 phrases. The average phrase length was about 520 ms. The professional also transcribed all of the utterances, and another one double-checked the annotated IPUs and transcripts.

Eye gaze. We used the facial-image processing tool OpenFace [36] to extract the eye gaze direction of participants from images of a camera that was placed in front of each participant. From the position of the camera and the seat of the participants, we calculated the participants' gaze direction toward the conversation partner. On the basis of gaze direction, we classified the participants' gaze targets: the head, body, and side of the head of the conversation partner, the lower half (the areas other than the participants below the participants' head), and the upper half (the area above the person's head).

Head direction. Using OpenFace on images of the participants obtained from the video camera placed in front of the participants, three-dimensional face orientation information, the angles of yaw, roll, and pitch, were acquired. We treated the head tilts of these three axes separately and divided each tilt by setting a threshold value. When each of these angles was 10 degrees or less, the tilt was labeled as center, 30 degrees or less as small (s), and over 30 degrees as large (l). However, in the upward direction of the head pitch (upper), there was no large amount of data, so we integrated the data classified as "upper".

Head nod. A head nod is a gesture in which the head tilts in alternating up and down arcs along the sagittal plane. A skilled annotator annotated the nods by observing upper body and overhead views in each frame of video. The annotated data was double-checked by another person. We regarded continuous nodding within a certain period as one nod event. The frequency (number) of nods was also manually labeled as 1, 2, 3, 4, 5 or more.

Hand gesture. Hand gestures being performed were manually annotated by one person and then double-checked by another person. The start of one hand gesture is when the hand starts moving from the home position. It also ends when the hand returns to the home position, or just before another kind of hand gesture is started. Hand gestures were classified into the following four major types based on McNeil's hand-gesture classification [28].

- Iconic: Gesture used to express scene depiction and motion.
- Metaphoric: This is a painterly and graphical gesture, but the content being depicted is abstract, for example, the flow of time.
- Beat: Used to represent the tone of utterances and emphasizing remarks such as by vibrating the hands and waving them in accordance with speech.
- Others: Gestures other than the above.

Upper-body posture. Participants were seated during the conversations, and there was no significant change how they were seated. For this reason, we extracted the forward and backward posture of the upper body on the basis of the three-dimensional position of the head. Specifically, we obtained the difference between the center position and coordinate position in the front-back direction of the head position obtained using OpenFace. On the basis of this positional information, the rotation of the upper body was calculated, and we classified upper-body postures into seven types of leaning posture. When the posture rotation was 10 degrees or less forward, it was classified as center, 15 degrees or less as micro small (forward_ss), 20 degrees or less as small (forward_s), 30 degrees or less as medium (forward_m), and over 45 degrees as large (forward_l). When it was 10 degrees or less backward, it was classified as center, 15 degrees or less, small (backward_s), and over 15 degrees, large (backward_l).

The temporal resolution of all verbal and nonverbal behavioral data was unified to 30 Hz.

3.2 Personality Traits

To assess the Big Five scores for each individual in the collected dialogues, 10 annotators were asked to carefully watch all conversation videos and answer all questions in a questionnaire [21]. Each dimension of the Big Five personality traits had a score from 1 to 7.

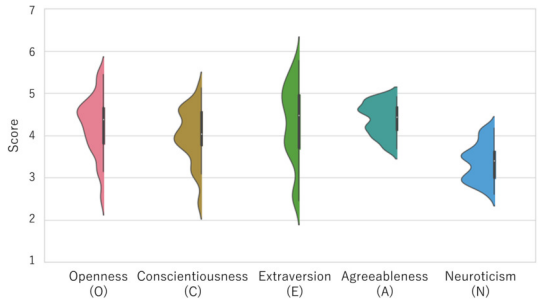

Figure 1: Distribution of Big Five scores of 22 participants

Figure 1 shows the distribution and box plot of each score of the Big Five for the 22 participants. Our goal is to generate the nonverbal behaviors of an agent from which users can perceive the personality traits of that agent. Although not all human personality traits are observable in conversational behaviors [12], for our research purposes, measuring the impression of a personality on the basis of third-party observation is more appropriate than using self-report scores. A similar method was used in [3].

In another piece of research [34], the authors collected personality trait scores from over 900 Japanese people and reported that the Extraversion score was widely distributed, but the distributions of Agreeableness and Neuroticism were much narrower. Our Big Five data in Figure 1 followed a very similar trend. We think that our data is a small sample that well reflects the tendency of the Big Five for Japanese people and is thus sufficient for our research.

We calculated the inter-rater agreement using the intraclass correlation coefficient (ICC). The average ICC score of all Big Five indicators was **0.668**: $ICC(2, 10) = \mathbf{0.362}$ for Openness (O), $ICC(2, 10) = \mathbf{0.748}$ for Conscientiousness (C), $ICC(2, 10) = \mathbf{0.648}$ for Extraversion (E), $ICC(2, 10) = \mathbf{0.836}$ for Agreeableness (A), and $ICC(2, 10) = \mathbf{0.748}$ for Neuroticism (N). The results suggest that the data were reliable. We used the average values of the 10 annotators as the participant's Big Five indicator scores.

4 EXPERIMENTAL SETUP

For analyzing the importance of Big Five information in nonverbal behavior, we implemented behavior generation models with/without Big Five information for each behavior type defined in Section 2.1 (Table 1). All models used utterance language information. In Section 4.1, we discuss the generation models and baselines that were used to perform sequential tagging of nonverbal behavior labels. We follow this up by detailing the implementation of our models in Section 4.2.

4.1 Generation Models

Our main goal is to study the impact of each personality trait in the Big Five vector on nonverbal behavior estimation. While numerous architectures can be employed as the generation model, our choice of a Bi-directional LSTM (or Bi-LSTM) [15] stems from its ability to model temporal dynamics from a smaller number of training samples. Furthermore, the two directions of processing can model both future and past contexts, hence providing the necessary information for gesticulation. The prediction of a sequence of nonverbal behavior labels is a structured prediction problem with

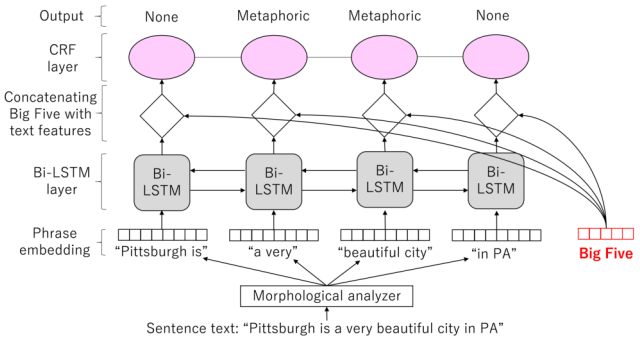

Figure 2: Our generation model using text and Big Five based on Bi-LSTM+CRF model

text as input and structured labels (such as the start and end of a nonverbal behavior label) as output. The prediction of structured output labels has been shown to improve with a conditional random field (CRF) layer on top of the Bi-LSTM [17, 26]. We used both personality and text information to estimate the energy function in a CRF layer that predicted the label for each nonverbal behavior type [20] in Table 1. A summary of this model is shown in Figure 2. These models generated one nonverbal behavior label for each phrase unit contained in the sentence. We extracted two kinds of text features for each phrase:

word2vec. We used a well-known pre-trained Japanese word2vec[1] using word information segmented by the morphological analyzer JUMAN++ [30], a Japanese morphological analyzer. We used the average vector of some 200 dimensional vectors of the word2vec features obtained from verbs, nouns and general adjectives included in phrases.

BERT. Spoken language was input to the BERT model for each phrase, and the average vector of the BERT features of all morphemes obtained from the final layer was used. We used one of the well-known pre-trained Japanese BERT models[2] using the spoken language information segmented by JUMAN++. With this method, a 768 dimensional vector was extracted for each phrase.

We implemented the following six prediction models. The first model always output the label with the highest number of samples in the training data (called "**Baseline**"), which was a strong baseline for unbalanced datasets. The second did not use text features but only used Big Five information (called "**Big5 model**"). By comparing this model and the baseline, we investigated the usefulness of nonverbal behavior generation with the Big Five alone. The next two models used only word2vec or BERT features (called "**word2vec model**" and "**BERT model**"). The final two models used Big Five along with word2vec or BERT features (called "**word2vec+Big5 model**" and "**BERT+Big5 model**").

4.2 Implementation Details

The sequential text (spoken language) features of word2vec or BERT, extracted from the phrases of the sentences, were adjusted to the maximum number of phrases in the sentences in the training set (which was 26 in our case). We used Keras [9] to train and test all models. We observed that training always converged to the best loss in the development set within five epochs or a wall time of around two hours. For our models, we used the sequential text features as an input to a Bi-LSTM layer with 50 units with the non-linear activation rectified linear unit (or ReLU) [1]. The 50-dimensional vectors extracted from the Bi-LSTM were then projected down to 20 dimensions by using a linear layer followed by ReLU. In parallel, a 5-dimensional Big Five vector was projected up to 20 dimensions by using a linear layer followed by ReLU to model the dynamics amongst the individual personality traits in the Big Five vector. Finally, these text and Big Five latent vectors were concatenated at each time step before feeding the vectors to the CRF layer for a structured prediction of the output labels. As the data has imbalanced classes, we set class sample weights to the inverse of the number of samples for each class in the training set [6].

[1]https://github.com/singletongue/WikiEntVec/releases
[2]http://nlp.ist.i.kyoto-u.ac.jp/index.php?BERT%E6%97%A5%E6%9C%AC%E8%AA%9EPretrained%E3%83%A2%E3%83%87%E3%83%AB

5 EXPERIMENT ON EFFECT OF BIG FIVE ON NONVERBAL BEHAVIOR GENERATION

5.1 Effect of Big Five

First, we evaluated if the Big Five was useful for generating individual nonverbal behaviors. To evaluate how well an individual person's nonverbal behavior labels could be estimated with the models using only the data of other persons, we used a 22-fold cross-validation (leave one-person out) technique with the data from the 22 participants.

We averaged the F-measures over all testing folds, and the results are shown in Table 2. By comparing the baseline and the models only using text features, the F-measure of the models using text features (word2vec and BERT models) were found to be significantly higher than the baseline (all results of paired t-test were $p < .01$). Therefore, text features such as the word2vec or BERT features were useful for generating actual individual whole nonverbal behaviors alone. By comparing the baseline and the models only using the Big Five (Big Five model), the F-measure of the Big Five model was found to be significantly higher than the baseline only for head direction (roll) and hand gesture (results of paired t-test: $t(21) = 4.82$, $p<.01$ for head direction (roll); $t(21) =$, $p<.01$ for hand gesture). Therefore, the Big Five features were useful for generating individual head rotation (roll) and hand gestures alone.

By comparing the models using only text features (word2vec and BERT models) and the model additionally using Big Five features (word2vec+Big5 and BERT+Big5 models), it was observed that the word2vec+Big5 model had significantly higher F-measures than the word2vec model did for only for the generation of eye gaze, hand gesture, and upper-body posture ($t(21) = 3.43$, $p<.01$ for eye gaze; $t(21) = 1.76$, $p<.10$ for hand gesture; $t(21) = 2.39$, $p<.05$ for upper body). The model using the Big Five (BERT+Big5 model) had statistically higher F-measures than the model without the Big Five (BERT model) for only the generation of eye gaze, head nod, and upper-body posture ($t(21) = 2.50$, $p < .01$ for eye gaze; $t(21) = 2.61$, $p<.05$ for head nod, for hand gesture; $t(21) = 6.39$, $p<.05$ for upper-body posture). Therefore, the Big Five information was useful for generating several individual nonverbal behavior labels such as eye gaze, head nod, hand gesture, and upper-body posture when using word and sentence representation.

Table 2: Mean and standard deviation of F-measure score for each type of nonverbal behavior generation model from 22-hold leave-one-person-out cross-validation. Results of paired t-test for three pairs of two conditions under Baseline vs Big5, word2vec vs word2vec+Big5, and BERT vs BERT+Big5 are shown in [ ] brackets. Results of paired t-test for each type of behavior under word2vec vs BERT+Big5 are shown in < > brackets. (: p<.01, *: p<.05, †: p<.10)**

	Baseline	Big5	word2vec	word2vec+Big5	BERT	BERT+Big5
Eye gaze	0.496 ± 0.145	0.510 ± 0.256	0.517 ± 0.236	**0.540** ± 0.235 [**]	0.531 ± 0.242	**0.552** ± 0.237 [*]
Head direction (yaw)	0.472 ± 0.169	0.478 ± 0.297	0.493 ± 0.275	0.491 ± 0.276	0.498 ± 0.273	0.491 ± 0.268
Head direction (roll)	0.607 ± 0.124	0.653 ± 0.274 [**]	0.672 ± 0.207	0.673 ± 0.207	0.671 ± 0.206	0.669 ± 0.204
Head direction (pitch)	0.430 ± 0.132	0.441 ± 0.249	0.444 ± 0.227	0.444 ± 0.226	0.444 ± 0.226	0.444 ± 0.226
Head nod	0.352 ± 0.046	0.377 ± 0.074	0.357 ± 0.071	0.395 ± 0.090	0.372 ± 0.071	**0.407** ± 0.071 [*]
Hand gesture	0.409 ± 0.104	0.488 ± 0.119 [**]	0.502 ± 0.137	**0.523** ± 0.125 [†]	0.506 ± 0.097	0.514 ± 0.137
Upper-body posture	0.281 ± 0.102	0.280 ± 0.222	0.265 ± 0.185	**0.321** ± 0.212 [*]	0.316 ± 0.229	**0.340** ± 0.235 [**],<†>

Table 3: Mean and standard deviation of F-measure score for each type of nonverbal behavior generation model from leave-one-person-out cross-validation. BERT+Big5 model was used for eye gaze, head nod, and upper-body posture generation. word2vec+Big5 model was used for hand-gesture generation.

	w/o O	w/o C	w/o E	w/o A	w/o N	w/ All (OCEAN)
Eye gaze (w/ BERT+Big5)	**0.529** ± 0.245 *	0.583 ± 0.240	0.582 ± 0.241	0.578 ± 0.238	0.581 ± 0.239	0.552 ± 0.237
Head nod (w/ BERT+Big5)	0.383 ± 0.064	0.391 ± 0.058	0.385 ± 0.051	**0.377** ± 0.072 †	0.376 ± 0.063	0.407 ± 0.707
Hand gesture (w/ word2vec+Big5)	**0.493** ± 0.123 *	**0.471** ± 0.136 †	0.529 ± 0.120	**0.493** ± 0.133 †	0.517 ± 0.119	0.523 ± 0.125
Upper-body posture (w/ BERT+Big5)	**0.284** ± 0.174 †	0.297 ± 0.199	0.292 ± 0.189	**0.294** ± 0.212 †	**0.274** ± 0.184 †	0.340 ± 0.235

We compared the performance of the word2vec+Big5 model and BERT+Big5 model to find the best performing model. As a result, the performance of the BERT+Big5 model for only upper-body posture generation was determined to have a trend that was significantly higher than that of the word2vec+Big5 model. This result suggests that BERT features are more useful than word2vec features for generating upper-body posture labels when using Big Five information ($t(21) = 1.82$, $p < .10$). When generating other nonverbal behavior labels, word2vec and BERT are no different in terms of their usefulness.

5.2 Effect of Each Big Five Indicator

The analysis in the previous section revealed that the Big Five helped generate nonverbal behaviors, such as eye gaze, head nod, hand gesture, and upper-body posture. Next, we verified which of the Big Five indicators is useful for generating which type of behavior. As a method for analysis, we constructed models that did not use one of the indicators from among the five indicators and compared the performance between them and models using all indicators. Since the BERT+Big5 model had the highest performance in terms of predicting eye gazes, head nods, and upper-body postures, and the word2vec+Big5 model had the highest performance in terms of predicting hand gestures, we used these models for this analysis. We constructed five models that generated each of the four nonverbal behavior labels, that is, for eye gazes, head nods, hand gestures, and upper-body postures. If a model that did not use an indicator performed statistically poorer than a model that used all indicators, it would be revealed that the unused indicator is useful for generating nonverbal behavior. The same method as in the previous section was used to build and evaluate the generation models.

The results are shown in Table 3. Regarding the performance in eye gaze generation, the model without the Openness indicator (w/o O) was significantly lower than the model using all indicators (w/ All) ($t(21) = 2.56$, p<.05). There was no difference between the models excluding the other indicators and the model using all indicators (w/ All). These results suggest that the Openness indicator is useful for generating individual eye gaze behavior.

Regarding the performance in head nod generation, the model without the Agreeableness indicator (w/o A) tended to be significantly lower than the model using all indicators (w/ All) ($t(21) = 1.76$, p<.10). This result suggests that the Agreeableness indicator is useful for generating individual head nod behavior.

For the performance in hand gesture generation, the models without the Openness, Agreeableness, and Conscientiousness indicators (w/o O, w/o A, and w/o C) were significantly lower than the model using all indicators (w/ All) ($t(21) = 2.14$, p<.05 for w/o O; $t(21) = 1.90$, p<.10 for w/o A; $t(21) = 1.76$, p<.10 for w/o C). This result suggests that the Openness, Agreeableness, and Conscientiousness indicators are useful for generating individual hand gesture behavior.

For the performance in upper-body posture generation, the models without the Openness, Agreeableness, and Conscientiousness indicators (w/o O, w/o A, and w/o C) were significantly lower than the model using all indicators (w/ All) ($t(21) = 1.95$, p<.10 for w/o O; $t(21) = 1.91$, $p < .10$ for w/o A; $t(21) = 2.02$, p<.10 for w/o C). This result suggests that the Openness, Agreeableness, and Conscientiousness indicators are useful for generating individual upper-body postures.

6 ANALYSIS ON RELATIONSHIP BETWEEN BIG FIVE INDICATORS AND NONVERBAL BEHAVIORS

In the previous section, we clarified which indicator has the potential to be useful in generating nonverbal behaviors. Finally, we analyzed the relationship between these indicator scores and nonverbal

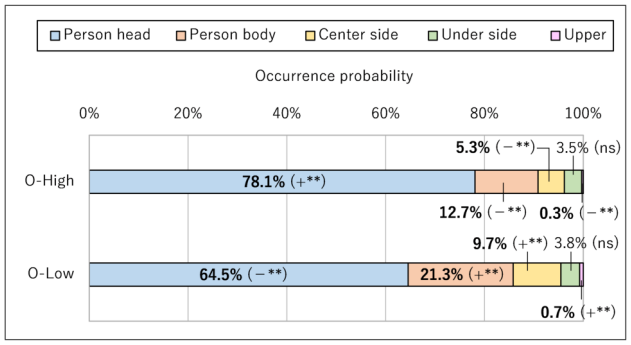

Figure 3: Occurrence probability of eye gaze labels for two person groups who have high or low Openness (O) score. (: p<.01, *: p<.05, †: p<.10)**

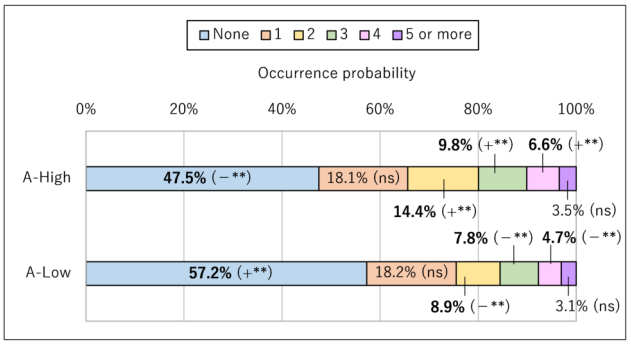

Figure 4: Occurrence probability of head nod label for two groups who had high or low Agreeableness (A) score.

behavior in detail. As a method for analysis, the 22 participants were divided into three groups of high (7 people), middle (8 people), and low (7 people) scores for each indicator. We compared the occurrence probability of nonverbal behavioral labels for phrases between the high and low groups, and this was done for each indicator that was useful for nonverbal behavior label generation in Section 5.

Relationship between Openness and eye gaze. Since the Openness (O) score is useful in eye gaze generation, we compared the occurrence probability of eye gaze labels between the high and low Openness score groups as shown in Figure 3. A chi-square was used to verify whether there was a difference in the probability between the two groups. As a result, there was significant difference between them ($\chi^2(4) = 300.4$, p<.01). In addition, a residual analysis was performed to verify which label was different between the two groups. The results are shown in Figure 3. The probability for "person head" was higher in the high score group than the low score group. Conversely, the probabilities for "person body," "center side," and "upper" were higher in the low score group than the high score group. These results suggest that those who had a high Openness score tended to look at a person's head more than those who had a low score.

Relationship between Agreeableness and Head Nod. Since the Agreeableness (A) score is useful in head nod generation, we compared the occurrence probability of head nod labels between the high and low Agreeableness score groups as shown in Figure 4.

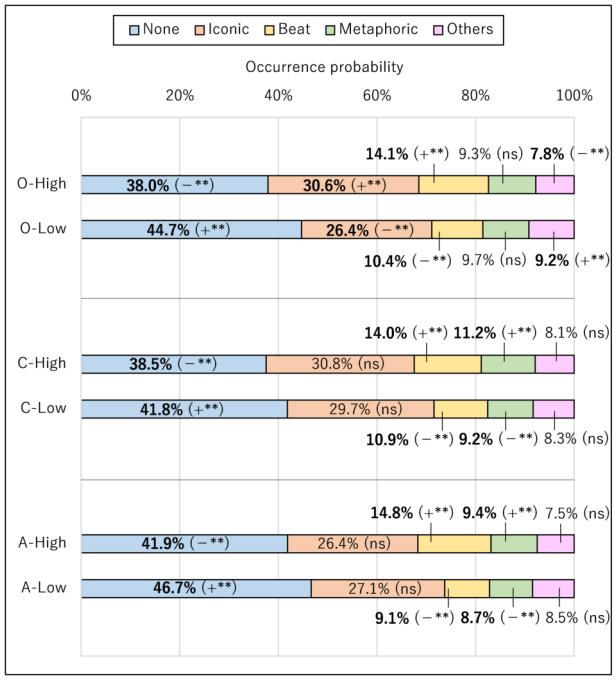

Figure 5: Occurrence probability of hand gesture label for two groups who had high or low scores for Openness (O), Conscientiousness (C), and Agreeableness (A)

There was significant difference in the probabilities between the two groups ($\chi^2(5) = 185.8$, p<.01). The result of a residual analysis is shown in Figure 4. The probability for "2," "3," and "4" (number of nods) was higher in the high score group than the low score group. Conversely, the probability for "none" was higher in the low score group than the high score group. These results suggest that those who had a high Agreeableness score tended to perform head nods two to four times more than those who had a low score.

Relationship between Openness, Conscientiousness, Agreeableness, and Hand Gesture. Since the Openness (O), Conscientiousness (C), and Agreeableness (A) scores are useful in hand-gesture generation, we compared the occurrence probability of hand gesture labels in each group of high and low Openness, Conscientiousness, and Agreeableness scores as shown in Figure 5. There were significant differences in the probabilities of hand gesture labels between the high and low Openness, Conscientiousness, and Agreeableness groups ($\chi^2(4) = 90.4$, p<.01 for O; $\chi^2(4) = 46.8$, p<.01 for C; $\chi^2(4) = 113.8$, p<.01 for A). The result of a residual analysis is shown in Figure 5.

The occurrence probabilities for "iconic" and "beat" were higher in the high score group for Openness (O) than the low score group. Conversely, the probabilities for none and others were higher in the low score group for Openness than the high score group. These results suggest that those who had a high Openness score tended to perform iconic and beat gestures more frequently than those who had a low score.

The occurrence probabilities for "beat" and "metaphoric" were higher in the high score group for Conscientiousness (C) than the low score group. Conversely, the occurrence probability for none

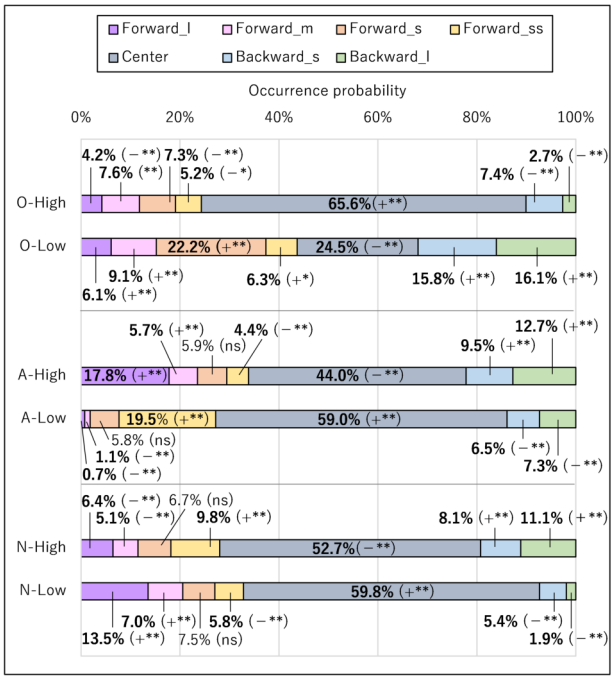

Figure 6: Occurrence probability of upper-body posture label for two groups who had high or low scores for Openness (O), Agreeableness (A), and Neuroticism (N)

was higher in the low score group than the high score group. These results suggest that those who had a high Conscientiousness score tended to perform beat and metaphoric gestures more than those who had a low score.

The occurrence probabilities for beat and metaphoric were higher in the high score group for Agreeableness (A) than the low score group. Conversely, the probabilities for none and others were higher in the low score group for Agreeableness than the high score group. These results suggest that those who had a high Agreeableness score tended to perform iconic and beat gestures more than those who had a low score.

Relationship between Openness, Agreeableness, and Neuroticism and Upper-body Posture. Since the Openness (O), Agreeableness (A), and Neuroticism (N) scores are useful in upper-body posture generation, we compared the occurrence probability of the upper-body posture labels in each group of high and low Openness, Agreeableness, and Neuroticism scores as shown in Figure 6. There were significant differences in the probabilities of the upper-body posture labels between the high and low groups for all three ($\chi^2(6) = 2425.0$, $p<.01$ for O; $\chi^2(6) = 2023.7$, $p<.01$ for C; $\chi^2(6) = 752.9$, $p<.01$ for A). The result of a residual analysis is shown in Figure 6.

The occurrence probabilities for "center" were higher in the high score group for Openness (O) than the low score group. Conversely, the probabilities of "forward_l," "forward_m," "forward_s," "forward_ss," "backward_s," and "backward_l" were higher in the low score group than the high score group. These results suggest that those who had high Openness scores tended to keep their upper-body posture in the center more than those who had a low score.

The occurrence probabilities of "forward_l," "forward_m," "backward_s," and "backward_l" were higher in the high score group for Agreeableness (A) than the low score group. Conversely, the probabilities of "forward_ss" and "center" were higher in the low score group for Agreeableness than the high score group. These results suggest that those who had high agreeableness scores tended to tilt the upper body backward and largely forward more than those who had a low score.

The occurrence probabilities of "forward_ss," "backward_s," and "backward_l" were higher in the high score group of Neuroticism (N) than the low score group. Conversely, the probabilities of "forward_l" and "forward_m" were higher in the low score group than the high score group. These results suggest that those who had high neuroticism scores tended to tilt the upper body backward and slightly forward more than those who had a low score.

7 DISCUSSION

The analysis of the results presented in Sections 5 and 6 suggests that the Big Five personality traits are useful for generating individual nonverbal behaviors on the basis of eye gaze, head nods, hand gestures, and upper-body postures. It was interesting that personality was not as useful for generating the head directions of yaw, roll, and pitch in our experiments. One possibility is that the differences may not have been noticeable since we created head labels from the average head position for each phrase. An interesting follow up analysis would be to study more detailed representations of head movements.

Previous research mainly demonstrated that there is a strong relationship between the Extraversion (E) indicator and nonverbal behaviors [10, 18, 24, 35]. It is interesting that we did not observe the same relationship with Extraversion. In our experimental results in Section 5.2, Extraversion was not useful in improving the prediction models for individual nonverbal behaviors. The reason could be due to cultural differences or a small amount of data.

We found interesting new trends with the other four indicators of Openness (O), Conscientiousness (C), Agreeableness (A), and Neuroticism (N). As an example, Agreeableness was associated with the most nonverbal behavior types, such as head nods, hand gestures, and upper-body postures. With higher Agreeableness, we observed more head nodding, iconic and beat gestures, and tilting of the upper body backward and largely forward. These results are consistent with the results of previous studies that analyzed the function of nonverbal behaviors. Head nodding is known to be used to approve other people's remarks and convey a positive message [37]. Regarding why more hand gestures and body postures occurred, making many hand gestures and leaning forward can be understood as the act of agreeing with and showing approval to a dialogue partner [29, 35]. Those who were high in Agreeableness seemed to have a tendency to do this.

We only dealt with Japanese dialogue data. Similar validation, using dialogue data from different languages and cultures, would be an interesting future research topic. We plan to analyze this in more detail with a larger dataset in the future since the ICC score for Openness (O) was the lowest (0.362) among all of the five indicators

in our corpus and Extraversion (E) was not useful for individual nonverbal behaviors generation.

Our generation model uses the five indicators of the Big Five as inputs simultaneously. Therefore, our model can combine the scores of the indicators to generate appropriate nonverbal behaviors. We have demonstrated the possibility of automatically generating full-body nonverbal behaviors that match an agent's personality from spoken language on the basis of the agent's Big Five personality traits set by the agent designer. In the future, we will collect a larger dataset and further improve our generation model. We will also evaluate whether users can perceive differences in an agent's personality from nonverbal behaviors in a similar way to [32, 38, 40].

8 CONCLUSION

In this paper, we demonstrated that the Big Five personality traits are useful when generating individual nonverbal behaviors including eye gaze, head nodding, hand gestures, and upper-body posture. In detail, the Openness (O) indicator may be useful for generating eye gaze, the Conscientiousness (C) indicator is useful for generating hand gestures, the Agreeableness (A) indicator is useful for generating head nods, hand gestures, and upper-body posture, and the Neuroticism (N) indicator is useful for generating upper-body posture. As future work, we would like to perform analysis the impact of personality traits for nonverbal behavior generation in more detail, especially Openness (O) and Extraversion (E), with a larger dataset. We will add speech information as input to our language-based nonverbal behavior generation model.

ACKNOWLEDGMENTS

This material is based upon work partially supported by the National Science Foundation (Awards 1722822, 1734868) and National Institutes of Health. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of National Science Foundation or National Institutes of Health, and no official endorsement should be inferred.

REFERENCES

[1] Abien Fred Agarap. Deep learning using rectified linear units (relu), 2018.
[2] Chaitanya Ahuja, Shugao Ma, Louis-Philippe Morency, and Yaser Sheikh. To react or not to react: End-to-end visual pose forecasting for personalized avatar during dyadic conversations. In *ICMI*, page 74–84, 2019.
[3] Oya Aran and Daniel Gatica-Perez. One of a kind: Inferring personality impressions in meetings. In *ICMI*, page 11–18, 2013.
[4] Jonas Beskow, Bjorn Granstrom, and David House. Visual correlates to prominence in several expressive modes. In *INTERSPEECH*, 2006.
[5] Carlos Busso, Zhigang Deng, Michael Grimm, Ulrich Neumann, and Shrikanth Narayanan. Rigid head motion in expressive speech animation: analysis and synthesis. In *Trans. ASLP*, pages 1075–1086, 2007.
[6] Jonathon Byrd and Zachary C Lipton. What is the effect of importance weighting in deep learning? *arXiv preprint arXiv:1812.03372*, 2018.
[7] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. Beat: The behavior expression animation toolkit. In *SIGGRAPH*, pages 477–486, 2001.
[8] Chung-Cheng Chiu, Louis-Philippe, and Stacy Marsella. Predicting co-verbal gestures: A deep and temporal modeling approach. In *ICMI*, volume 9238, pages 152–166, 08 2015.
[9] François Chollet et al. Keras. https://keras.io, 2015.
[10] Kevin Frank. *Posture & Perception in the Context of the Tonic Function Model of Structural Integration: An Introduction.* IASI Yearbook, 2007.
[11] Takeshi Fuchi and Shinichiro Takagi. Japanese morphological analyzer using word co-occurrence -JTAG-. In *COLING*, pages 409–413, 1998.
[12] Robert Gifford. ch. personality and nonverbal behavior: a complex conundrum. *The SAGE Handbook of Nonverbal Communication*, pages 159–181, 2006.
[13] Shiry Ginosar, Amir Bar, Gefen Kohavi, Caroline Chan, Andrew Owens, and Jitendra Malik. Learning individual styles of conversational gesture. In *CVPR*, pages 3497–3506, 2019.
[14] Hans Peter Graf, Eric Cosatto, Volker Strom, and Fu Jie Huang. Visual prosody: Facial movements accompanying speech. In *FG*, pages 381–386, 2002.
[15] Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. Bidirectional LSTM networks for improved phoneme classification and recognition. In *ICANN*, page 799–804, 2005.
[16] Dai Hasegawa, Naoshi Kaneko, Shinichi Shirakawa, Hiroshi Sakuta, and Kazuhiko Sumi. Evaluation of Speech-to-Gesture generation using Bi-Directional LSTM network. In *IVA*, page 79–86, 2018.
[17] Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional LSTM-CRF models for sequence tagging. *ArXiv*, abs/1508.01991, 2015.
[18] Katherine Isbister and Clifford Nass. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *J. Human-Computer Studies*, 53(2):251–267, 2000.
[19] Ryo Ishii, Taichi Katayama, Ryuichiro Higashinaka, and Junji Tomita. Automatic generation system of virtual agent's motion using natural language. In *IVA*, pages 357–358, 2018.
[20] Ryo Ishii, Taichi Katayama, Ryuichiro Higashinaka, and Junji Tomita. Generating body motions using spoken language in dialogue. In *IVA*, pages 87–92, 2018.
[21] Oliver P. John. The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. 1990.
[22] Yuki Kadono, Yutaka Takase, and Yukiko I. Nakano. Generating iconic gestures based on graphic data analysis and clustering. In *HRI*, pages 447–448, 2016.
[23] Munhall KG, Jeffery A Jones, Daniel E Callan, Takaaki Kuratate, and Eric Vatikiotis-Bateson. Visual prosody and speech intelligibility: Head movement improves auditory speech perception. In *Psychol Sci.*, volume 15, pages 133–137, 2004.
[24] Mark L Knapp and Gerald R Miller. Communicator characteristics and behavior. *Handbook of Interpersonal Communication*, pages 103–161, 1994.
[25] Taras Kucherenko, Dai Hasegawa, Gustav Eje Henter, Naoshi Kaneko, and Hedvig Kjellström. Analyzing input and output representations for speech-driven gesture generation. In *IVA*, page 97–104, 2019.
[26] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. In *NAACL*, pages 260–270, 2016.
[27] Manja Lohse, Reinier Rothuis, Jorge Gallego-Pérez, Daphne E. Karreman, and Vanessa Evers. Robot gestures make difficult tasks easier: The impact of gestures on perceived workload and task performance. In *CHI*, pages 1459–1466, 2014.
[28] David McNeill. *Hand and Mind: What Gestures Reveal About Thought.* University Of Chicago Press, 1996.
[29] Albert Mehrabian. Significance of posture and position in the communication of attitude and status relationships. *Psychological Bulletin*, 71(5):359–372, 1969.
[30] Hajime Morita, Daisuke Kawahara, and Sadao Kurohashi. Morphological analysis for unsegmented languages using recurrent neural network language model. In *EMNLP*, pages 2292–2297, 2015.
[31] Yukiko I. Nakano, Masashi Okamoto, Daisuke Kawahara, Qing Li, and Toyoaki Nishida. Converting text into agent animations: assigning gestures to text. In *NAACL*, pages 153–156, 2004.
[32] Michael Neff, Yingying Wang, Rob Abbott, and Marilyn Walker. Evaluating the effect of gesture and language on personality perception in conversational agents. In *IVA*, pages 222–235, 2010.
[33] Fumio Nihei, Yukiko I. Nakano, Ryuichiro Higashinaka, and Ryo Ishii. Determining iconic gesture forms based on entity image representation. In *ICMI*, pages 419–425. ACM, 2019.
[34] Atsushi Oshio, Shingo Abe, and Pino Cutron. Development, reliability, and validity of the japanese version of ten item personality inventory (TIPI-J). *Jap. J. Personality*, 21(40–52), 2012.
[35] RichardLippa. The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: a lens model analysis. *J. Research in Personality*, 32(1):80–107, 1998.
[36] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: a unified embedding for face recognition and clustering. *CoRR*, 2015.
[37] Senko Maynard. Japanese conversation: Self-contextualization through structure and interactional management. *Ablex Publishing Corporation*, 1989.
[38] Harrison Jesse Smith and Michael Neff. Understanding the impact of animated gesture performance on personality perceptions. *ACM Trans. Graph.*, 36(4), July 2017.
[39] Petra Wagner, Zofia Malisz, and Stefan Kopp. Gesture and speech in interaction: An overview. 57:209–232, 2014.
[40] Yingying Wang, Jean E. Fox Tree, Marilyn Walker, and Michael Neff. Assessing the impact of hand motion on virtual character personality. *ACM Trans. Appl. Percept.*, 13(2), March 2016.
[41] Hani Camille Yehia, Takaaki Kuratate, and Eric Vatikiotis-Bateson. Linking facial animation, head motion and speech acoustics. 30(3):555–568, 2002.