UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
PersonalityScanner: Exploring the Validity of Personality Assessment Based on Multimodal Signals in Virtual Reality

Permalink
https://escholarship.org/uc/item/4p8191z5

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Zhang, Xintong
Lu, Di
Hu, Huiqi
et al.

Publication Date
2024

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

PersonalityScanner: Exploring the Validity of Personality Assessment Based on Multimodal Signals in Virtual Reality

Xintong Zhang [1,2], Di Lu [3], Huiqi Hu [1], Nan Jiang [2,4], Xianhao Yu [1], Jinan Xu [1], Yujia Peng [2,3,4], Qing Li[2], Wenjuan Han[✉1]

[1]Beijing Key Lab of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing, China
[2]State Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence (BIGAI)
[3]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University
[4]Institute for Artificial Intelligence, Peking University

Abstract

Human cognition significantly influences expressed behavior and is intrinsically tied to authentic personality traits. Personality assessment plays a pivotal role in various fields, including psychology, education, social media, *etc*. However, traditional self-report questionnaires can only provide data based on what individuals are willing and able to disclose, thereby lacking objective. Moreover, automated measurements and peer assessments demand significant human effort and resources. In this paper, given the advantages of the Virtual Reality (VR) technique, we develop a VR simulator — `PersonalityScanner`, to stimulate cognitive processes and simulate daily behaviors based on an immersive and interactive simulation environment, in which participants carry out a battery of engaging tasks that formulate a natural story of first-day at work. Through this simulator, we collect a synchronous multi-modal dataset with ten modalities, including first/third-person video, audio, text, eye tracking, facial microexpression, pose, depth data, log, and inertial measurement unit. By systematically examining the contributions of different modalities on revealing personality, we demonstrate the superior performance and effectiveness of `PersonalityScanner`.

Keywords: Personality Assessment; Virtual Environment; Automatic Assessment

The cognitive process encompasses perception, memory, thinking, judgment, and problem-solving. These psychological processes influence how individuals interpret and understand the world around them, thereby affecting their responses and behaviors (Kogan & Wallach, 1964; Schaie, Willis, & Caskie, 2004), including body posture, facial expressions, and voice. For instance, an optimistic cognitive orientation may lead a person to take proactive actions when facing challenges. The mode of cognition affects the choice of behavior, and these behavioral patterns accumulate over time, forming the unique personality traits of an individual. Personality refers to the enduring patterns of behavior, emotion, and thought processes of an individual and has been studied extensively in fields of psychology, education, and social media, *etc*. (Costa Jr & McCrae, 2008; Kalat, 2016). Personality assessments help individuals better understand their own strengths, weaknesses, motivations, and preferences. This self-awareness can be beneficial in a variety of areas, such as making better decisions, career planning, relationship building, and personal development (Caspi, Roberts, & Shiner, 2005; Weiner & Greene, 2017).

Traditional personality assessments primarily rely on self-reporting of personality traits (Weiner & Greene, 2017) and peer evaluations (McNeill, 2000; Oosterhof & Todorov, 2008). However, the traditional approaches can only yield information that people are able and willing to report rather than real responses controlled by their cognition and suffer from a lack of objective, multi-dimensional, and automatic measurements. To overcome these limitations, the field of computational personality assessment has emerged (W. Ilmini & Fernando, 2017; Palmero et al., 2021) leveraging computational techniques to predict personality traits more comprehensively and objectively. Among computational approaches, virtual reality (VR) based approaches recently attracted widespread interest (Majumder, Poria, Gelbukh, & Cambria, 2017; Iacobelli, Gill, Nowson, & Oberlander, 2011), due to potential advantages of VR (*e.g.*, experimental design' ease of creation, control, and standardization (Maharani, 2023), safe environment with minimal exposure to hazardous situations (Maharani, 2023), effectively recording multi-modal sensor information (Xie et al., 2021)). For instance, (Katifori, Lougiakis, & Roussou, 2022) explored how personality traits influence participants' performance of object manipulation tasks in VR, specifically the task of selecting and moving a cube inside a hole. They proposed several related hypotheses for verification, such as participants with higher levels of openness are more likely to exhibit increased hand movements. Similarly, (Maharani, 2023) asked participants to perform asphalt compaction tasks by manually steering the wheel and joystick to reveal the relationship between personality and specific task performance. However, both studies focus solely on basic object manipulation tasks and only consider hand movements, neglecting other body movements and interactions with others, which can provide valuable insights (Sackett, Gruys, & Ellingson, 1998).

In this work, we aim to utilize VR technology to identify human cognitive-controlled behaviors during the execution of complex daily tasks to reveal personality traits. To the best of our knowledge, no prior research has explored the connection between personality traits and complex daily-life activities within an interactive VR simulator that allows participants to freely move their whole bodies, engage in open-ended conversations with non-playable characters (NPCs), and interact with the environment. To address this gap, we develop `PersonalityScanner`, an interactive VR simulator with advanced NPCs, personality-sensitive tasks, and versatile tech-

✉ Corresponding author.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

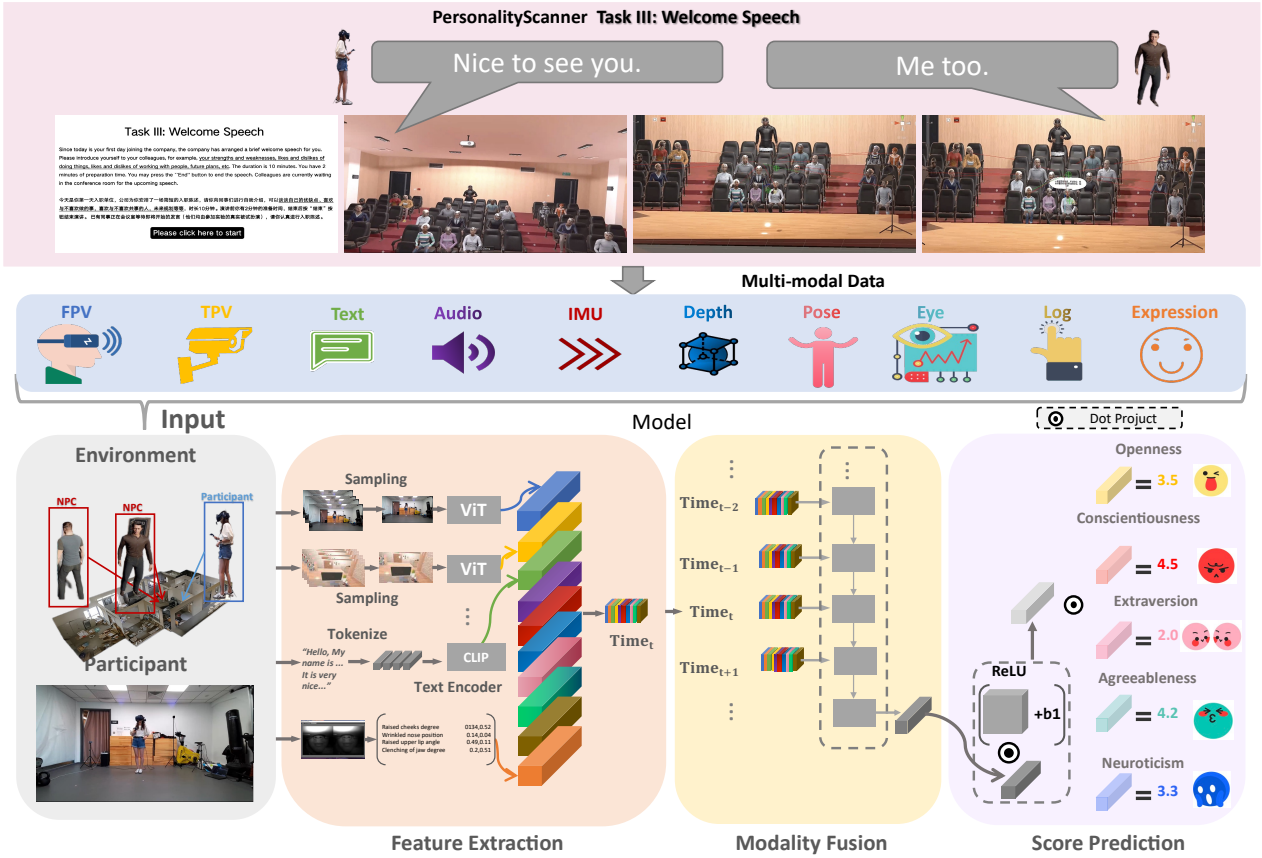

Figure 1: A overview of `PersonalityScanner` simulator and our model to estimate personality.

nological features including real-time data processing, high frame rates, modular VR control, and user-friendly 3D map editing. In `PersonalityScanner`, participants experience their first day at a company, interact with NPCs, and complete tasks (as seen in Figure 1). We collect multi-modal data during the experiments, which we later use to train a model that maps behavioral performance to personality traits. This model enables us to predict participants' personalities when enough data is available.

In summary, our study makes three significant contributions: (1) We introduce `PersonalityScanner`, an interactive and immersive VR simulator, capable of simulating everyday cognitive environments and social interactions. (2) We collect a synchronous multimodal dataset using the VR simulator, encompassing ten modalities. This dataset serves as a new benchmark for evaluating personality assessment performance. (3) Our comprehensive study of different modality types proves the superiority of `PersonalityScanner` and our collected dataset in estimating personality traits.

Related Work

Human Cognition and Personality Traits

According to research findings (Kumari, Williams, Gray, et al., 2004), personality traits such as extraversion, conscientiousness, and neuroticism are closely associated with human brain structure and gray matter activity, which serve as critical determinants and controllers of human cognitive processes. Previous studies in psychology have also frequently asserted that individuals' personalities are intimately connected with their cognitive processes in various everyday activities, such as adventurous behaviors (Kogan & Wallach, 1964) and creativity (McCrae, 1993). Importantly, (Schaie et al., 2004) shows that the link between human cognition and personality is stable, sometimes lasting over 35 years. This suggests that cognition is a reliable indicator of personality.

Figure 2: Three tasks of `PersonalityScanner` simulator.

The widely-used Big Five Personality Factors (Big

Five) (Costa & McCrae, 1999; McCrae & Costa Jr, 1997) is used as our theoretical foundation for assessing personality. It categorizes human personality using five key traits:

- **Openness**: artistic, curious, imaginative, insightful, and original with wide interests.
- **Conscientiousness**: efficient, organized, planful, reliable, responsible, and thorough.
- **Extraversion**: active, assertive, energetic, enthusiastic, outgoing, and talkative.
- **Agreeableness**: appreciative, forgiving, generous, kind, and sympathetic.
- **Neuroticism**: anxious, self-pitying, tense, touchy, unstable, and worrying.

In our research, we use the International Personality Item Pool (IPIP) and its IPIP-NEO derivations (Goldberg et al., 1999, 2006; Johnson, 2005, 2014), along with the BFI-S questionnaire by (Lang, John, Lüdtke, Schupp, & Wagner, 2011) to provide the initial personality labels for participants. Each item in the questionnaire consists of a question and a set of options. All items are labeled with the corresponding Big Five personality factors annotated by psychologists for standardized personality assessment.

Personality Traits Assessment

Personality traits are crucial as they reflect long-term behavioral patterns with impacts on many life aspects (Caspi et al., 2005). Studies show these traits significantly influence key life outcomes (Weiner & Greene, 2017; Caspi et al., 2005), emphasizing the need for regular personality assessments in quality-of-life surveys. Researchers have developed a variety of computational methods, employing diverse machine learning algorithms, to assess personality traits effectively.

Self-reported Personality Traits Assessment. Researchers have investigated whether self-reported personality traits can be evaluated accurately, *etc*. For example, (Wolffhechel et al., 2014) utilized the Cubiks In-depth Personality Questionnaire (CIPQ 2.0), a normative self-reported tool, to measure various personality traits.

Peer Assessment. There is reliable psychological and biological evidence (Kumari et al., 2004) suggesting that multimodal behaviors serve as reliable predictors of personality traits. These include audio signals (Fang, Achard, & Dubuisson, 2016), visual cues such as facial expressions and gestures (Salam, Celiktutan, Hupont, Gunes, & Chetouani, 2016), and observable social cues (Principi, Palmero, Junior, & Escalera, 2019).

Computational Personality Traits Assessment. Humans can assess some personality traits from features, prompting research into whether computers can similarly automatically evaluate them. For instance, face-image based approaches (K. Ilmini & Fernando, 2016; QIN, GAO, XU, & HU, 2018), audio based approaches (Gürpınar, Kaya, & Salah, 2016; Zhang, Zhang, Wei, & Wu, 2016), and face-to-face dyadic interaction-based scenarios (Palmero et al., 2021).

Personality Traits in Virtual Reality

The importance of personality traits as factors significantly affecting the user experience in a virtual environment (VE) has been recognized in previous research, leading to several studies attempting to explore these effects (Majumder et al., 2017; Iacobelli et al., 2011). There has been research conducted on the effects of specific personality traits in VR environments, with studies dating back as far as 20 years. The personality traits most frequently studied in VR environments include absorption (Tellegen & Atkinson, 1974), mental imagination (Sheehan, 1967), immersive tendencies (Qin, Patrick Rau, & Salvendy, 2009), locus of control (Rotter, 1966), empathy (Davis, 2018). While this research highlights the need for further exploration of personality traits and their impact on the user experience in VR environments, such existing studies have primarily focused on presence (Wallach, Safir, & Samana, 2010; Kober & Neuper, 2013) and, to a lesser extent, on other aspects of the VR experience, such as the sense of embodiment(Dewez et al., 2019).

Recent studies by (Katifori et al., 2022) and (Maharani, 2023) focus on task performance in VR and explore predicting user personality traits. Their research was limited to simple object manipulation tasks using hand movements. To our knowledge, There is no known research on the link between personality traits and performance in everyday activities in VR settings.

`PersonalityScanner` Simulator

Functions

We develop an interactive and immersive VR simulator to stimulate cognitive progress, referred to as `PersonalityScanner`. This simulator is equipped with a range of advanced functionalities as follows.

Daily Life Task Design. The primary goal of our simulator is to immerse users in everyday tasks that can reveal their personality traits. Since narrative elements can help to draw the users into the VR experience and make them more engaging, we integrate a role-playing narrative comprising a series of challenges. Participants engage in a "company onboarding" scenario, including riding an elevator with strangers, organizing a desk, and delivering a welcome speech, as depicted in Figure 2. These tasks were selected because of their relevance to human personality traits. In the elevator task, different personalities lead to varying choices in standing position, behavior in crowded elevators, and gaze direction (Fryrear, Lane, & Itzkowitz, 1976; Rousi & Renko, 2020). Desk organization correlates with the "conscientiousness" trait in the Big Five personality model (Freeman, 1990), reflecting the degree of personal item orderliness. During public speaking, individuals with distinct personalities exhibit different performance indicators, including eye movements (Kim et al., 2018) and vocal responses, when they are observed by an audience, revealing their unique psychological states (Kuai, Liang, He, & Wu, 2021).

Specifically, we instruct each participant to imagine themselves as a new employee about to join a multinational technology company and complete the onboarding process on the first day of employment with the company. The participant's initial task is to take an elevator ride with unknown colleagues and head to the office. Subsequently, they should gather office supplies and organize their desk. The final task involves going to the conference room to deliver a brief welcome speech to their colleagues. The instructions are presented to the participant at the beginning of each task.

Interactive NPC. Believable NPCs of human behavior can enhance interactive applications across a variety of immersive environments (Park et al., 2023). To develop VR applications that are both immersive and engaging, we design NPCs capable of giving natural feedback to the user's pose, touch, and voice commands. This design is based on multimodal large language models (MLLMs) that have been proven to understand human languages (Vaswani et al., 2017; Wei et al., 2023) and provide reactions similar to persons' personality (Jiang et al., 2022), aiming to make interactions as natural and intuitive as possible, thereby enriching the overall VR experience. The NPCs display natural body movements like fidgeting or gesturing during conversations, instead of staying static. They also have spontaneous conversations with each other. When touched by the user, NPCs turn to face them naturally. They act and respond like colleagues during discussions. If the participant is visible or talking to them, NPCs maintain eye contact by following the user's movements.

Unanticipated Event Generation. According to existing research (Farizi, Bangay, & Mckenzie, 2019; Craig, 2013), movements within VR environments tend to be involuntary, thus making conscious control of these movements a challenging endeavor. Micro-expressions are often regarded as indicators of truthfulness since individuals usually cannot suppress their spontaneous behaviors in situations that impose cognitive load or involve responding to unanticipated questions (DePaulo et al., 2003).

Fortunately, unanticipated events can be easily created by `PersonalityScanner`. Take the third task *Organizing Your Desk* as an example, we design a mechanism that can occasionally and automatically disrupt the items on the office table without warning. Based on our experimental observations, participants with high conscientiousness tend to make repeated efforts to organize items that have been disrupted. On the other hand, participants who do not have high conscientiousness but try to pretend to be conscientious are more likely to give up on tidying up.

Objects. The environment in which our personality assessment occurs can greatly impact the level of immersion. Consequently, we strive to create a visually appealing and interactive environment. For instance, objects within this environment can be grasped and moved. The elevator button can be pressed. When pressed, lights up to signal its activation, and the door opens.

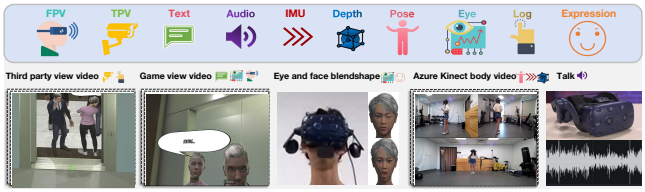

Figure 3: Illustration of collected dataset and modalities.

Usage

The recorded data includes multimodal information from participants who wear a head-mounted depth sensor, and a spatial microphone array enacting their roles in a simulated household environment, as shown in Figure 3 and Figure 2. Our experimental site is equipped with Azure Kinect DK, an integrated device featuring a depth sensor, spatial microphone array, video camera, and orientation sensor, complemented by comprehensive software development kits (SDKs).

The usage steps of `PersonalityScanner` are as follows:

Prepare the VR environment. The VR environment needs to be properly set up. Subsequently, the participant is required to wear the VR headset and be able to move around freely in the environment.

Introduce the experiment to participants. Each participant signs an informed consent form, and is informed of an introduction to the experiment, details of the experimental procedures, information about potential risks, a privacy statement, compensation details, and a 10-minute exercise.

Start the assessment. Each participant is required to engage in three tasks, each lasting 10 minutes. Throughout the assessment, two managers will oversee the participants to ensure their comfort and safety. Additionally, three psychology experts will observe the participants' behaviors and performance, and they will validate and revise the self-assessed answers to the OCEAN personality test completed by the participants at the time of registration.

Equipment Participants are required to complete three tasks while wearing a VR headset, a facial expression tracker, and holding handsets as depicted in Figure 2. We utilize the Vive Pro headset and Azure Kinect, which includes an RGB camera, depth sensor, spatial microphone array with a video camera, and orientation sensor as an all-in-one device with software development kits. As for the platform, we leverage the Unity game engine to develop immersive VR scenarios and execute these applications on the SteamVR platform.

Dataset

Task Formulation

We introduce a new benchmark for personality prediction by recording the multimodal data of selected participants as they complete tasks within `PersonalityScanner`.

Participants Selection

To recruit suitable participants, recruitment posters are disseminated through social media platforms and online forums,

requiring each prospective participant to provide basic personal information, including gender and age, *etc*. And complete an OCEAN personality traits questionnaire. When selecting participants, we place particular emphasis on balancing the gender ratio and personality traits among participants. We ensure a 1:1 male-to-female ratio and select participants based on their OCEAN personality questionnaire results, aiming to balance the representation of each trait within the participants. Furthermore, the study mainly involves individuals aged 18 to 35, who tend to adapt to VR technology quickly, which improves the quality of our dataset.

Data Construction

We collect data from the selected participants, and three psychology experts review and revise the initial personality test through the observation of the participant's performance during interactions. The dataset after verification and correction is then utilized for training the personality prediction model.

Data Collection. The dataset comprises multimodal information recorded while participants perform tasks in `PersonalityScanner`, as shown in Figure 3. We collect data from 150 participants, including ten modalities such as first/third-person video, audio, text, eye tracking, facial microexpression, pose, depth data, log, and inertial measurement unit. Additionally, it includes information about positions, interaction with objects and NPCs, *etc*. The human poses in each frame of the dataset are labeled following the COCO human pose style.

Quality Control in Expert Review and Labeling. Considering the preliminary questionnaire results are based on participants' self-reports and reflect their subjective views of their personality traits, they are not completely accurate. To improve the dataset's reliability and accuracy, we have three psychology experts observe the participants during tasks and adjust their questionnaire results if needed. The final personality assessments are decided by majority vote among these experts, ensuring objectivity and accuracy in our dataset.

Dataset Analysis. The detailed statistics of our collected dataset are shown in Table 1. The dataset comprises RGB+D video frames, the mesh of objects scene point clouds, action labels, and activity labels. This dataset encompasses data from 150 participants, consisting of 450 videos, 24.3 million RGB-D frames, and 20 motion categories. The total duration of recordings for all participants is over 4500 minutes, with a total number of 27000 × 5 clips (5 for the following video types including RGB video, depth image, audio, skeleton, and bounding box). We draw the bounding box of the object that the participant interacts with, along with hand localization. The datasets have been segmented into 10-second clips, and divided into into training, development, and testing with a 0.6/0.2/0.2 ratio.

Table 1: Statistics of multi-modal dataset.

Modality	Amount
RGB+D Body video	450
Multi-view video	300
Eye and face blendshape	150
Audio	150
Text	150
Supplementary log	600

Personality Traits Inference

This section assesses the collected dataset's capacity to infer personality traits. We utilize a multi-modal transformer-based model (Vaswani et al., 2017) to predict the personality traits of participants in `PersonalityScanner`.

In our research, we predict personality traits by translating human activities in a simulator into five OCEAN scores, treating each as a separate regression task. We use a transformer-based model to estimate these traits, converting observed behaviors within `PersonalityScanner` into quantifiable OCEAN personality scores.

The model comprises three modules: feature extraction, modality fusion, and OCEAN score prediction. The lower part of Figure 1 provides a visual representation of all components and their interconnecting information flow. Detailed descriptions of these components are also presented below.

- **Feature extraction**: Human activities generate synchronous multi-modal data that can be recorded by `PersonalityScanner`. Following (Zhao et al., 2023), different sources of context are captured in the form of different modalities including first/third-person view video, audio, text, eye tracking, facial microexpression, pose, depth data, Logs, and Inertial Measurement Unit (IMU). In the feature extraction module, we employ separate encoders for each modality. We leverage the CLIP model (Radford et al., 2021) for image and text encoder. We convert audio modality into 2D mel-spectrograms (Gong, Chung, & Glass, 2021), and depth modalities into one-channel images and adopt ViT-S (Tong, Song, Wang, & Wang, 2022) encoders respectively. The body joint coordinates are obtained offline from Azure Kinect camera videos, and we address occlusion issues by calibrating the cameras and selecting unobstructed frames for data extraction. These joint coordinates are then encoded using a Multilayer Perceptron (MLP). For face and eye blend shapes, Temporal Convolutional Networks (TCN) (Lea, Flynn, Vidal, Reiter, & Hager, 2017) are utilized.
- **Modality fusion**: To obtain a fixed-size embedding that is both normalized and utilized in the InfoNCE loss (Oord, Li, & Vinyals, 2018), we incorporate a linear projection head specific to each modality into every encoder. All sensors are synchronized. We utilize the attention mechanism to model the relation between each pair of two modalities to obtain the ultimate input for the transformer layers.
- **OCEAN score prediction** We obtain OCEAN traits by applying a fully connected layer of the transformer. An Ocean score is computed for each participant.

The assessment of results is carried out based on the Mean Squared Error (MSE) between the aggregated personality trait score and the corresponding ground truth label (*i.e*.,

Expert-reviewed and verified OCEAN personality test results.) for each participant in the test set.

Experiments and Results

Dataset Comparison We compare the datasets we collected with relevant personality research datasets. To the best of our knowledge, our dataset is the first to be collected within a VR environment, encompassing ten modalities for a comprehensive assessment of personality, setting it apart from existing datasets. See Table 2 for a numerical comparison.

The Cocktail Party dataset (Zen, Lepri, Ricci, & Lanz, 2010) is a pioneering collection for recording six-person social interactions to analyze spatial and attention dynamics related to personality. CoffeeBreak (Cristani et al., 2011) examines social group structures and orientations. IDIAP Poster (Hung & Kröse, 2011) offers overhead views for detecting F-Formations during presentations. Big Game (Hung, Englebienne, & Kools, 2013) focuses on identifying social actions in a quiz game context with sensor and video data. SALSA (Alameda-Pineda et al., 2015) integrates video, sensor information, and personality traits for social behavior analysis. MatchNMingle (Cabrera-Quiros, Demetriou, Gedik, van der Meij, & Hung, 2018) documents interactions in speed-dating and parties with comprehensive behavioral annotations. UDIVA (Palmero et al., 2021) provides a non-acted dyadic interaction dataset with audiovisual, physiological, and sociodemographic data, along with personality and relationship insights.

Table 2: Comparison of Personality research datasets.

Dataset	Numb. of people	Total time (minutes)	VR	Modalities				
				V	A	F	P	I
Cocktail	6	30	✗	✓	✗	✗	✗	✗
CoffeeBreak	10	-	✗	✓	✗	✗	✗	✗
Big Game	32	30	✗	✓	✗	✗	✗	✗
Idiap	50	360	✗	✓	✗	✗	✗	✗
SALSA	18	60	✗	✓	✓	✗	✗	✓
MatchNMingle	92	120	✗	✓	✓	✗	✗	✓
UDIVA	147	**5430**	✗	✓	✓	✗	✗	✓
Ours	**150**	4500	✓	✓	✓	✓	✓	✓

Table 3: Accuracy results of estimating the personality traits. The higher, the better.

Methods	O	C	E	A	N	All
FGM-UTRECHT	70.36	65.22	73.98	60.46	56.14	65.23
SMART-SAIR	72.65	**67.02**	74.52	64.93	57.32	67.29
Ours	**75.43**	66.96	**75.93**	**68.03**	**60.87**	**69.44**

Baselines Due to the scarcity of work utilizing multimodal data for personality prediction, we compare our method with SMART-SAIR (Salam, Celiktutan, Mukherjee, Manoranjan, & Ismail, 2021) and FGM-UTRECHT (Pessanha & Sogancioglu, 2022), two works in the ICCV 2021 Understanding Social Behavior Challenge (Palmero et al., 2021). Building on studies about gender-specific personality traits (Weisberg, DeYoung, & Hirsh, 2011), SMART-SAIR separates participants by gender and crafted a gender-specific multimodal model using Neural Architecture Search (Jin, Song, & Hu, 2019), integrating visual and textual data for personality prediction. FGM UTRECHT develops a Random Forest (RF) regressor using only metadata features like age, gender, and session count. They also combined linguistic, audio, and metadata features in a multitask approach using RF regressors and late fusion.

Estimate Personality Traits Using `PersonalityScanner`. We compare our method with SMART-SAIR and FGM-UTRECHT. As shown in Tables 4 and Table 3. From Table 4, Notably, our approach achieved the lowest mean square error for the Openness score, followed by the Extraversion score, outperforming the two baselines. This suggests that our method provides better estimates of self-assessed personality traits. Table 3 presents a summary of the accuracy comparison.

Table 4: MSE results of estimating the personality traits. The lower, the better.

Methods	O	C	E	A	N	All
FGM-UTRECHT	0.9640	0.9147	0.9852	0.9039	1.0033	0.9542
SMART-SAIR	0.8434	**0.8302**	0.8894	0.8823	0.9071	0.8705
Ours	**0.8112**	0.8554	**0.8298**	**0.8801**	**0.8839**	**0.8521**

Impact of Modalities. We conduct an ablation study to demonstrate the influence of various modalities. As shown in Table 5, the removal of modalities results in a further reduction of prediction capability. These findings underscore the significant role that different modalities play.

Table 5: Accuracy of estimating personality traits. We analyze the impact of different modalities. The higher, the better. A-Audio, T-Text, P-Pose, F-Face Blendshape, I-Interaction.

Modality	O	C	E	A	N	All
A	58.23	54.02	56.96	57.39	52.27	55.77
T	56.83	52.96	50.49	57.93	56.02	54.85
P	54.20	53.87	55.77	53.20	52.86	53.98
F	55.79	56.04	58.31	53.98	54.28	55.68
A-T	65.39	61.10	66.99	62.83	58.24	62.91
A-T-P	67.29	63.89	70.65	64.20	60.98	65.40
A-T-P-F	70.33	64.71	72.73	65.43	**61.26**	66.89
A-T-P-F-I	**75.43**	**66.96**	**75.93**	**68.03**	60.87	**69.44**

Conclusion

We develop a VR simulator to immerse the participant in the VR experience and make it more engaging to reveal their personality. With this simulator, we collect a large-scale, synchronous, multi-modal dataset with ten modalities. Furthermore, we conduct extensive research on different types of modalities and demonstrate the superiority of `PersonalityScanner`.

Acknowledgments

We express our gratitude to the anonymous reviewers for their valuable comments. Special thanks to BIGAI for providing equipment and support for the experimental site, and heartfelt appreciation to every participant who signed up. The work is supported by the National Science and Technology Major Project of China (2022ZD0114900), the Talent Fund of Beijing Jiaotong University (2023XKRC006) to WH, the National Natural Science Foundation of China (32200854) and the Young Elite Scientists Sponsorship Program (2021QNRC00) to YP.

References

Alameda-Pineda, X., Staiano, J., Subramanian, R., Batrinca, L., Ricci, E., Lepri, B., … Sebe, N. (2015). Salsa: A novel dataset for multimodal group behavior analysis. *IEEE transactions on pattern analysis and machine intelligence*, *38*(8), 1707–1720.

Cabrera-Quiros, L., Demetriou, A., Gedik, E., van der Meij, L., & Hung, H. (2018). The matchnmingle dataset: a novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates. *IEEE Transactions on Affective Computing*, *12*(1), 113–130.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annu. Rev. Psychol.*, *56*, 453–484.

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. In *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510–516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, P. T., & McCrae, R. R. (1999). A five-factor theory of personality. In *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). The Guilford Press New York, NY, USA.

Costa Jr, P. T., & McCrae, R. R. (2008). *The revised neo personality inventory (neo-pi-r).* Sage Publications, Inc.

Craig, C. (2013). Understanding perception and action in sport: how can virtual reality technology help? *Sports Technology*, *6*(4), 161–169.

Cristani, M., Bazzani, L., Paggetti, G., Fossati, A., Tosato, D., Del Bue, A., … Murino, V. (2011). Social interaction discovery by statistical analysis of f-formations. In *Bmvc* (Vol. 2, pp. 10–5244).

Davis, M. H. (2018). *Empathy: A social psychological approach*. Routledge.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, *129*(1), 74.

Dewez, D., Fribourg, R., Argelaguet, F., Hoyet, L., Mestre, D., Slater, M., & Lécuyer, A. (2019). Influence of personality traits and body awareness on the sense of embodiment in virtual reality. In *2019 ieee international symposium on mixed and augmented reality (ismar)* (pp. 123–134).

Fang, S., Achard, C., & Dubuisson, S. (2016). Personality classification and behaviour interpretation: An approach based on feature categories. In *Proceedings of the 18th acm international conference on multimodal interaction* (pp. 225–232).

Farizi, F. D., Bangay, S., & Mckenzie, S. (2019). Facial cues for deception detection in virtual reality based communication. In *Proceedings of the 3rd international conference on big data and internet of things* (pp. 65–69).

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill.

Freeman, W. J. (1990). Consciousness as physiological self-organizing process. *Behavioral and Brain Sciences*, *13*(4), 604–605.

Fryrear, J. L., Lane, E. J., & Itzkowitz, J. R. (1976). Who presses the elevator button after someone else: Dominance/deference, sex, and authority figure determinants. *Psychological reports*, *38*(2), 671–674.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96.

Goldberg, L. R., et al. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, *7*(1), 7–28.

Gong, Y., Chung, Y.-A., & Glass, J. (2021). Ast: Audio spectrogram transformer. *arXiv preprint arXiv:2104.01778*.

Gürpınar, F., Kaya, H., & Salah, A. A. (2016). Combining deep facial and ambient features for first impression estimation. In *Computer vision–eccv 2016 workshops: Amsterdam, the netherlands, october 8-10 and 15-16, 2016, proceedings, part iii 14* (pp. 372–385).

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year old. *Cognition and Brain Theory*, *6*, 287–317.

Hung, H., Englebienne, G., & Kools, J. (2013). Classifying social actions with a single accelerometer. In *Proceedings of the 2013 acm international joint conference on pervasive and ubiquitous computing* (pp. 207–210).

Hung, H., & Kröse, B. (2011). Detecting f-formations as dominant sets. In *Proceedings of the 13th international conference on multimodal interfaces* (pp. 231–238).

Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). Large scale personality classification of bloggers. In *International conference on affective computing and intelligent interaction.*

Ilmini, K., & Fernando, T. (2016). Persons' personality traits recognition using machine learning algorithms and image processing techniques. *Advances in Computer Science: an International Journal*, *5*(1), 40–44.

Ilmini, W., & Fernando, T. (2017). Computational personality traits assessment: A review. In *2017 ieee international con-*

ference on industrial and information systems (iciis) (pp. 1–6).

Jiang, G., Xu, M., Zhu, S.-C., Han, W., Zhang, C., & Zhu, Y. (2022). Mpi: Evaluating and inducing personality in pre-trained language models. *arXiv preprint arXiv:2206.07550*.

Jin, H., Song, Q., & Hu, X. (2019). Auto-keras: An efficient neural architecture search system. In *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining* (pp. 1946–1956).

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129.

Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the ipip-neo-120. *Journal of Research in Personality*, *51*, 78–89.

Kalat, J. W. (2016). *Introduction to psychology*. Cengage Learning.

Katifori, A., Lougiakis, C., & Roussou, M. (2022). Exploring the effect of personality traits in vr interaction: The emergent role of perspective-taking in task performance. *Frontiers in Virtual Reality*, *3*, 19.

Kim, H., Shin, J. E., Hong, Y.-J., Shin, Y.-B., Shin, Y. S., Han, K., … Choi, S.-H. (2018). Aversive eye gaze during a speech in virtual environment in patients with social anxiety disorder. *Australian & New Zealand Journal of Psychiatry*, *52*(3), 279–285.

Kober, S. E., & Neuper, C. (2013). Personality and presence in virtual reality: Does their relationship depend on the used presence measure? *International Journal of Human-Computer Interaction*, *29*(1), 13–25.

Kogan, N., & Wallach, M. A. (1964). Risk taking: A study in cognition and personality.

Kuai, S.-G., Liang, Q., He, Y.-Y., & Wu, H.-N. (2021). Higher anxiety rating does not mean poor speech performance: dissociation of the neural mechanisms of anticipation and delivery of public speaking. *Brain Imaging and Behavior*, *15*(4), 1934–1943.

Kumari, V., Williams, S. C., Gray, J. A., et al. (2004). Personality predicts brain responses to cognitive demands. *Journal of neuroscience*, *24*(47), 10636–10641.

Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the big five: Robust across survey methods except telephone interviewing. *Behavior Research Methods*, *43*(2), 548–567.

Lea, C., Flynn, M. D., Vidal, R., Reiter, A., & Hager, G. D. (2017). Temporal convolutional networks for action segmentation and detection. In *proceedings of the ieee conference on computer vision and pattern recognition* (pp. 156–165).

Maharani, A. (2023). *The relationship between personality traits and task performance as a function of complexity*. Unpublished master's thesis, University of Twente.

Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, *32*(2), 74–79.

Matlock, T. (2001). *How real is fictive motion?* Doctoral dissertation, Psychology Department, University of California, Santa Cruz.

McCrae, R. R. (1993). Openness to experience as a basic dimension of personality. *Imagination, Cognition and Personality*, *13*(1), 39–55.

McCrae, R. R., & Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509.

McNeill, D. (2000). *The face: a natural history*. Back Bay Books.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Ohlsson, S., & Langley, P. (1985). *Identifying solution paths in cognitive diagnosis* (Tech. Rep. No. CMU-RI-TR-85-2). Pittsburgh, PA: Carnegie Mellon University, The Robotics Institute.

Oord, A. v. d., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092.

Palmero, C., Selva, J., Smeureanu, S., Junior, J., Jacques, C., Clapés, A., … others (2021). Context-aware personality inference in dyadic scenarios: Introducing the udiva dataset. In *Proceedings of the ieee/cvf winter conference on applications of computer vision* (pp. 1–12).

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442*.

Pessanha, F., & Sogancioglu, G. (2022). Fact sheet: Automatic self-reported personality recognition track. *arXiv preprint arXiv:2207.11012*.

Principi, R. D. P., Palmero, C., Junior, J. C. J., & Escalera, S. (2019). On the effect of observed subject biases in apparent personality analysis from audio-visual signals. *IEEE Transactions on Affective Computing*, *12*(3), 607–621.

Qin, H., Patrick Rau, P.-L., & Salvendy, G. (2009). Measuring player immersion in the computer game narrative. *Intl. Journal of Human–Computer Interaction*, *25*(2), 107–133.

QIN, R., GAO, W., XU, H., & HU, Z. (2018). Modern physiognomy: an investigation on predicting personality traits and intelligence from the human face. *Information Sciences*, *61*(058105), 1–058105.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., … others (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748–8763).

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological*

monographs: General and applied, *80*(1), 1.

Rousi, R., & Renko, R. (2020). Emotions toward cognitive enhancement technologies and the body–attitudes and willingness to use. *International Journal of Human-Computer Studies*, *143*, 102472.

Sackett, P. R., Gruys, M. L., & Ellingson, J. E. (1998). Ability–personality interactions when predicting job performance. *Journal of applied psychology*, *83*(4), 545.

Salam, H., Celiktutan, O., Hupont, I., Gunes, H., & Chetouani, M. (2016). Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access*, *5*, 705–721.

Salam, H., Celiktutan, O., Mukherjee, H., Manoranjan, V., & Ismail, I. (2021). Fact sheet: Automatic self-reported personality recognition track. In *Iccv 2021 understanding social behavior in dyadic and small group interactions challenge.*

Schaie, K. W., Willis, S. L., & Caskie, G. I. (2004). The seattle longitudinal study: Relationship between personality and cognition. *Aging Neuropsychology and Cognition*, *11*(2-3), 304–324.

Sheehan, P. W. (1967). A shortened form of betts' questionnaire upon mental imagery. *Journal of clinical psychology*.

Shrager, J., & Langley, P. (Eds.). (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences (" absorption"), a trait related to hypnotic susceptibility. *Journal of abnormal psychology*, *83*(3), 268.

Tong, Z., Song, Y., Wang, J., & Wang, L. (2022). Videomae: Masked autoencoders are data-efficient learners for self-supervised video pre-training. *Advances in neural information processing systems*, *35*, 10078–10093.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wallach, H. S., Safir, M. P., & Samana, R. (2010). Personality variables and presence. *Virtual Reality*, *14*, 3–13.

Wei, X., Cui, X., Cheng, N., Wang, X., Zhang, X., Huang, S., … others (2023). Zero-shot information extraction via chatting with chatgpt. *arXiv preprint arXiv:2302.10205*.

Weiner, I. B., & Greene, R. L. (2017). *Handbook of personality assessment*. John Wiley & Sons.

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the big five. *Frontiers in psychology*, *2*, 178.

Wolffhechel, K., Fagertun, J., Jacobsen, U. P., Majewski, W., Hemmingsen, A. S., Larsen, C. L., … Jarmer, H. (2014). Interpretation of appearance: The effect of facial features on first impressions and personality. *PloS one*, *9*(9), e107721.

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., … others (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality*, *2*, 645153.

Zen, G., Lepri, B., Ricci, E., & Lanz, O. (2010). Space speaks: towards socially and personality aware visual surveillance. In *Proceedings of the 1st acm international workshop on multimodal pervasive video analysis* (pp. 37–42).

Zhang, C.-L., Zhang, H., Wei, X.-S., & Wu, J. (2016). Deep bimodal regression for apparent personality analysis. In *European conference on computer vision* (pp. 311–324).

Zhao, H., Cai, Z., Si, S., Ma, X., An, K., Chen, L., … Chang, B. (2023). Mmicl: Empowering vision-language model with multi-modal in-context learning. In *The twelfth international conference on learning representations.*