ORIGINAL ARTICLE

VIRTUAL HUMANS FOR ASD INTERVENTION: A BRIEF SCOPING REVIEW

LUNA MADDALON[1], **MARIA ELEONORA MINISS**I[1], **SERGIO C. TORRES**[1],
SOLEDAD GÓMEZ-GARCÍA[2], **MARIANO ALCAÑIZ**[1]

[1]*Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano (HUMAN-tech), Universitat Politécnica de Valencia, Valencia, España.* [2]*Facultad de Magisterio y Ciencias de la Educación, Universidad Católica de Valencia, Valencia, España*

Abstract   Individuals with autism spectrum disorder may present social-communicative and behavioral deficits. Recently, research on treatment and diagnosis has shifted its focus to the application of new technologies. Among them is virtual reality, which guarantees a high sense of realism to the experience and allows the implementation of a virtual agent that facilitates the use of the application. In social skills interventions, it has been mostly chosen to implement a virtual agent with a human appearance. Virtual humans guide the user-system interaction through the use of verbal and nonverbal language. They can be equipped with responsiveness: the ability to provide responses to the user based on data recorded during the use of the technology. Responsiveness is functional when the goal is to create an interaction similar to that of everyday life, as it allows for behavioral responses and, at a more sophisticated level, vocal responses. Considering virtual agents capable of holding a conversation with the user, to date three different methods have been implemented that make communication more or less realistic. This brief review proposes a synopsis of relevant virtual humans' features and highlights some key ASD research areas wherein virtual humans are implemented for diagnosis and treatment. A total of 11 studies were selected and their analysis was summarized into 7 main categories. Finally, the clinical and technological implications of the results found were discussed.

Key words: virtual human, conversational agent, virtual agent, autistic spectrum disorder, training, intervention

Resumen   *Humanos virtuales para la intervención en TEA: una breve revisión del alcance*
    Los individuos con trastorno del espectro autista pueden presentar déficits socio-comunicativos y conductuales. Recientemente, la investigación sobre el tratamiento y el diagnóstico se ha centrado en la aplicación de nuevas tecnologías. Entre ellas se encuentra la realidad virtual, que garantiza un alto sentido de realismo a la experiencia y permite la implementación de un agente virtual que facilite el uso de la aplicación. En las intervenciones de habilidades sociales, se ha optado mayoritariamente por implementar un agente virtual con apariencia humana. Los humanos virtuales guían la interacción usuario-sistema mediante el uso de lenguaje verbal y no verbal. Estos pueden estar dotados de responsividad: la capacidad de proporcionar respuestas al usuario basadas en los datos registrados durante el uso de la tecnología. La responsividad es funcional cuando el objetivo es crear una interacción similar a la de la vida cotidiana, ya que permite dar respuestas conductuales y, a un nivel más sofisticado, respuestas vocales. Considerando los agentes virtuales capaces de mantener una conversación con el usuario, hasta la fecha se han implementado tres métodos diferentes que hacen que la comunicación sea más o menos realista. Esta breve revisión propone una sinopsis de las características de los humanos virtuales relevantes y destaca algunas áreas de investigación clave del TEA en las que se implementan humanos virtuales para el diagnóstico y el tratamiento. Se seleccionó un total de 11 estudios y su análisis se resumió en 7 categorías principales. Por último, se discuten las implicaciones clínicas y tecnológicas de los resultados encontrados.

Palabras clave: humano virtual, agente conversacional, agente virtual, trastorno del espectro autista, entrenamiento, intervención

Autism spectrum disorder (ASD) is a neurodevelopmental disorder affecting approximately1 out of 100 children worldwide[1]. Nuclear symptoms refer to persistent deficits in communication and social interaction, and restrictive and repetitive behavioral patterns, interests,

or activities[2]. Besides standard methodology for ASD diagnosis (semi-structured interviews and questionnaires), there is an increasing interest in implementing techniques capable of providing objective assessments based on biometrics. In recent decades, due to the heterogeneous nature of the disorder, the interest in using implicit measures for the detection of objective biomarkers in ASD has been growing[3]. The advantage of these techniques lies in the possibility of capturing the explicit

Postal address: Luna Maddalon - Univ. Politécnica Valencia. I3B/CPI cubo 8B - Camino Vera s/n, 46022, Valencia, España
e-mail: lmaddal@i3b.upv.es

and implicit manifestations of the disorder. Moreover, these objective measures are helpful to design tailored interventions addressing individualized needs. Some of these techniques include electroencephalography, eye tracking, video analysis, video modeling, voice analysis, and event-related potentials[3], which serve to "sensorize" the technological system. In other words, the technological system is capable of recording and analyzing users' psychophysiological responses (in real-time or offline) to subsequently guide the training/intervention. In this way, it can be defined as an intervention "system equipped with sensorization".

ASD research is increasingly trending toward the implementation of new technologies over traditional methodologies. Several studies choose to adopt standard computer-based or mobile applications in the ASD field due totheir greater objectivity in stimuli administration[4]. Elsewhere, robotic technologies have been opted to strengthen the presence of an agent or character capable of communicating and moving in space[4]. A meta-analysis on the topic provided evidence that the most trained ability in ASD using new technologies is social skills[5]. Specifically, initiating a conversation, social conventions, responding to others, nonverbal behaviors, regulating emotions and reciprocity, and relationships, are key research targets. In addition to the use of mobile applications and robots, virtual reality (VR) has been suggested as an effective tool for the assessment and training of individuals with ASD[6].The strength of VR relies on its safeness in providing controlled and realistic content, which enables transferring learning effects to real situations[6].VR could be an economical alternative to the robot, considering the advantage of being able to convey richer and more ecological human-like interactions through the inclusion of virtual agents.

A virtual agent is a computer program that can take any form and appearance depending on how it is coded. Virtual humans exhibit human-like behaviors, such as speech, gestures, and movements; at a finer layer, they can get to exhibit emotions, planning, motivation, and memory[7]. By supervising the verbal and nonverbal communicative behaviors of virtual humans, besides ensuring a good level of realism, it is possible to induce socially relevant cognitive processes in the user with whom they interact[8].This sort of virtual agent with human-like appearances enables realistic human interactions in the virtual environment. Therefore, it supposes a very useful mean for analyzing social dynamics, which are associated with one of the most common ASD deficits[8]. Several types of deployments of these agents are possible. In particular, virtual humans can be included in virtual environments in a passive way by guiding the virtual experience (through standardized instructions, questions, or feedback), or in an active way, through responsiveness, which is the quality of being responsive. Virtual humans endowed with responsive-

ness have the ability to modify their verbal or nonverbal information with regard to the information gathered from the user. On a technical level, the user's expressions (e.g., semantic utterances, prosodic elements, body movements) are processed and analyzed by computer programs as inputs, and the response (verbal or not) to yield is proposed and formulated by the virtual human. To ensure the accuracy and quality of the responsive feedback, it is necessary to train the computer model through multiple inputs and possible responses. Owing this feature, virtual human can be considered adaptive to different situations generated by the users. The application of responsive virtual humans can be leveraged to provide new directions for ASD therapy, given the potential capabilities to make treatments highly individualized[4]. Indeed, virtual humans can communicate using non-verbal cues, such as gesture, pointing, or eye gaze, and linguistic features, such as prosody cues, written text, and speech[7].The latter allows the creation of a context similar to natural human language, due to the responsiveness they possess.

Shifting the focus to virtual humans capable of communicating through language-based modalities, two control modules have been mostly used to date. Language-based control modules allow the virtual agent to generate conversations based on the manner in which they are programmed. Response-retrieval dialogue method is based on a predefined and recorded set of verbal or written information that the system retrieves in response to user's input[9].On the other hand, the response-generation method relies the response on the probability distribution learned from the training data, consistent with the user's input[9].

This article aimed at providing an overview of the studies that have adopted a virtual agent having human-like appearances in the treatment and evaluation of ASD. The goal was to investigate the characteristics and capabilities that these virtual humans hold (particularly if they are able to converse), how they operate on the user, and to which purpose (i.e., assessment, intervention or training).

Materials and methods

The literature search was carried out through the Scopus electronic database using the following keywords: TITLE-ABS-KEY("virtual human" OR "conversational agent" OR "virtual character" OR "avatar") AND ("autism" OR "autistic spectrum disorder" OR "ASD") AND ("diagnosis" OR "assessment" OR "training" OR "intervention" OR "clinical*").Inclusion criteria addressed: a) research articles and conference articles published during the last 10 years, b) ASD participants between 6 and 19 years old (or with an average of less than 19 y. o.), c) virtual agent with human features (e.g., eyes, mouth, etc.), d) technology-based intervention, training or comprehensive assessment on ASD, and e) virtual agent's core feature focusing on training or evaluating the user. Exclusion criteria addressed: a) systematic reviews, book chapters, or meta-analyses, b) articles published in non-English languages, c) articles not reporting the age and number of participants, and d) articles not fully available.

Findings have been categorized into: virtual human interface, virtual human appearance, system sensorization, virtual human responsivity, interaction module, trained ability, and participants' age. The first category has been split into two possible interfaces of the virtual human: 2D or 3D. Regarding the appearance of the virtual human, it meant that it can be depicted only at the face and torso level, full-body, or in both modes. A sensorization category was included to address all those possible implicit (i.e., psychophysiological) and explicit measures (i.e., behavioral response, semantics) regarding the user's inputs that were recorded by the system during the experience. The virtual human responsivity referred to the degree to which it is able to be responsive to users and to adapt to their inputs. It has been intended that through the implicit and explicit data collected by the system, the virtual human may or may not be responsive. The interaction module category reported the extent to which the virtual human is able to naturally interact with the user. There were two options leveraging the degree of interactions: predetermined sentences (i.e., automatic and standardized sentences delivered at certain moments of the experience without considering the user's input) when the virtual human did not naturally interact, and response generation method. Finally, the core ability investigated by the application tested in the study and the participant's age was reported.

Results

Eleven out of eighty-eight papers accomplished the inclusion criteria. The eleven selected articles were presented in Figure 1.

The studies involved 184 participants in total (of whom 33 -18%- were children or teenagers with typical development). Eight articles implemented a 2D virtual human; the remaining used a 3D representation. Six studies focused on presenting the upper part of the virtual human (face and trunk) two studies focused on the full body. Finally, three more studies implemented both modalities, and in two studies, full body. Regarding the degree of system sensorization, the different studies addressed several approaches: behavioral response (5), facial expressions (5), head orientation (3), semantics (3), prosody (3), eye movements (2), and motor movements (1). Among the eleven articles, only six implemented adaptive virtual humans characterized by responsivity. The degree of virtual human responsivity varied among the selected articles which involved facial expression (3), head orientation (2), eye gaze (2), audiovisual prompts (1), finger pointing (1), and head nods (1). Focusing on those virtual humans capable of producing speech content, four of them deployed a type of speech interaction that did not follow natural conversational criteria as it utilized predetermined

Fig.1-— Selected articles presented depending on the publication year

Article	Virtual agent interface	Virtual agent appearance	System sensorization	Virtual agent responsivity	Interaction module	Trained ability	Age
Jyoti V, Lahiri U. (2022)	2D	Full body	Facial orientation, behavioral response	Finger-pointing, head orientation	Predetermined sentences	Social skills	6 to 12
Santos P, et al. (2022)	2D	Both	Facial expressions, facial orientation	Facial expressions	/	Social skills	6 to 8
Genova HM, et al. (2021)	2D	Upper part	Behavioral response	/	Predetermined sentences	Job interview	16 to 19
Amat AZ, et al. (2021)	2D	Upper part	Eye movements	Eye gaze, head orientation, behavioral response	/	Social skills	7 to 13
Ali MR, et al. (2020)	2D	Upper part	Facial expressions, semantics, prosody	/	Response generation	Social skills	13 to 17
Milne M, et al. (2018)	2D	Upper part	Behavioral response	/	Predetermined sentences	Social skills	6 to 12
Tanaka H, et al. (2017)	2D	Upper part	Facial orientation, facial expressions, semantics, prosody	Head nods	Predetermined sentences	Social skills	7 to 19
Bekele E, et al. (2016)	3D	Both	Facial expressions, eye movements	Facial expressions, eye gaze	Response generation	Social skills	13 to 17
Razavi SZ, et al. (2016)	2D	Upper part	Facial expressions, semantics, prosody	/	Response generation	Social skills	15 to 17
Serret S, et al. (2014)	3D	Both	Behavioral response	/	/	Social skills	6 to 17

sentences. Three articles instead were programmed by the response generation dialogue method, thus capable of holding a conversation. Nine articles aimed to train social skills, one aimed to train hand-eye coordination, and one to train for job interviews referring to social integration. Finally, the participants' mean age was under 13 y. o in four cases, and greater than 13 y.o. in seven cases.

Discussion

The objective of this article was to review recent studies introducing virtual agents with human-like features (i.e., virtual humans) for ASD treatment or diagnosis. Specifically, the most important properties in terms of system sensorization and agent features were summarized. The focus on virtual technologies stemmed from the notion that they offered effective methods for treating ASD deficits; in particular, the implementation of virtual humans in VR can be useful for both verbal and nonverbal individuals due to the enhanced feeling of realistic presence in social interactions[8].

Regarding the interface and appearance of the virtual human in the selected articles, the analysis opened an important issue of this research trend, which is the confusion in the definition of specific technical terms. For instance, this limitation occurred in Genova et al. (2021), in which the definition of "virtual human" is placed in the context of Virtual Reality Job Interview Training (VR-JIT). Other literature stated that VR-JIT systems did not use drawn virtual humans but used videos of real people[8] instead. It would be appropriate, therefore, to yield a consensus on the conceptual definition of what a virtual human is and whether and how it differs from other terms such as virtual agent or avatar. In this paper, a potential definition of what is a virtual human is proposed.

The results proved that most of the studies choose to sensorize the developed system to monitor behavioral responses, facial expressions, head orientation, semantics, prosody, and eye movements[10-19]. This is consistent with the investigated ability, mostly social skills, and with the type of responsiveness of the virtual human. The degree of system sensorization, functional to the responsive adaptation of virtual humans, tended to work in six articles with real-time techniques. On the contrary, in the rest of the articles, the virtual human did not adapt in real-time the response depending on user's biometric inputs. The real-time adaptation allowed response patterns to vary based on the input (behavioral or physical) data acquired, and to return individualized output in real-time[4]. The peculiarity of virtual humans lied in their ability to modify the information they transmit based on the user's recorded data. The degree of

virtual human responsivity varied among the selected articles depending on the acquired information. In this regard, by modifying their own facial, head, and gaze characteristics, such virtual humans were able to send prompts to the user[10, 11, 13, 16, 17, 20]. For example, a study trained joint attention in ASD individuals through a game-based system in which the virtual human adapted the gaze and head orientation with the aim of improving the user's performance[13]. Another example was the case of "Zeca", a virtual human that was used with the goal of having the user copy the virtual humans' facial expressions[11]. The basic idea was that through a model by imitation, ASD individuals could train facial movements, expressions, and identify emotions.

Regarding the interaction module category, virtual humans presented in the articles can be divided into three categories: 1) those that did not present conversational content[11, 13, 19, 20], 2) those that have implemented predetermined sentences that are reproduced automatically regardless of user's input[10, 12, 15, 16], and 3) those that were programmed via artificial intelligence (AI) able to modulate their speech according to the training data (i.e., response generation method)[14, 17, 18]. The latter approach is the most sophisticated of those presented so far in this paper. Response generation refers to a dialogue-oriented chat capturing implicit and/or explicit user data with the aim of generating a coherent response through probability calculations[9]. Response generation is closely related to natural language processing (NLP), which is the field of AI dealing with human language processing. For example, response generation can use NLP to analyze the structure and meaning of sentences in the input and use this information to generate a coherent and relevant response. In short, response generation depends on the NLP to function properly. Technologies that use machine learning algorithms to generate consistent responses include: chat bots, virtual assistance systems, and automatic translation systems.

Future research should implement more sophisticated software version like, for example, Generative Pre-trained Transformer (GPT). GPT is a language model developed by Open AI using response generation that employs deep learning techniques to analyze and understand human language to parse the input and generate an appropriate response. Besides being more modern and stable, GPT, and particularly GPT-3, which is the latest version, guarantees a high sense of realism in the natural language of the virtual agent. A further option would be the Megatron-Turing Natural Language Generation model (MT-NLG), but it is not open source as GPT-3. The implementation of this type of software in a virtual human presented in VR would ensure a sense of realism and ecological validity in the treatment and diagnosis of ASD.

TAKE-HOME MESSAGE

Current knowledge

1. Responsive virtual humans adapt in real time their output according to the user input gathered.

2. Two language-based modalities have been mostly used to date: the response-retrieval and response-generation methods.

Contribution of the article

1. Overview of empirical evidence for the use of conversational responsive agents in ASD research.

2. Future directions within this research area have been outlined as scientific advances.

Bibliography

1. World Health Organization [WHO]. Autism Spectrum Disorders (2019). Available at: *https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders.*
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed., text rev.), 2022.
3. Bölte S, Bartl-Pokorny KD, Jonsson U, et al. How can clinicians detect and treat autism early? Methodological trends of technology use in research. *Acta paediatrica* 2016; 105: 137-44.
4. Alcañiz M, Maddalon L, Minissi ME, et al. Intervenciones tecnológicas adaptativas para el trastorno del espectro autista: una revisión bibliográfica. *Medicina (B Aires)* 2022; 82: 54-8.
5. Grynszpan O, Weiss PL, Perez-Diaz F, et al. Innovative technology-based interventions for autism spectrum disorders: a meta-analysis. *Autism* 2014; 18: 346-61.
6. Parsons S. Authenticity in Virtual Reality for assessment and intervention in autism: A conceptual review. *Educational Research Review* 2016; 19: 138-57.
7. Burden D, Savin-Baden M. Virtual humans: Today and tomorrow. Chapman and Hall/CRC, 2019.
8. Lugrin B, Pelachaud C, Traum D. (Eds.). The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, Volume 2: Interactivity, Platforms, Application, 2022.
9. Lubis N, Sakti S, Yoshino K, et al. Positive emotion elicitation in chat-based dialogue systems. IEEE/ACM. *Transactions on Audio, Speech, and Language Processing* 2019; 27: 866-77.
10. Jyoti V, Lahiri U. Portable joint attention skill training platform for children with autism. *IEEE Transactions on Learning Technologies* 2022; 15: 290-300.
11. Santos P, Silva V, Sena-Esteves J, et al. HiZeca: A Serious Game for Emotions Recognition. In International Conference Innovation in Engineering. Springer, Cham. 2021; 393-405.
12. Genova HM, Lancaster K, Morecraft J, et al. A pilot RCT of virtual reality job interview training in transition-age youth on the autism spectrum. *Research in Autism Spectrum Disorders* 2021; 89: 101878.
13. Amat AZ, Zhao H, Swanson A, et al. Design of an interactive virtual reality system, In: ViRS, for joint attention practice in autistic children. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2021; 29: 1866-76.
14. Ali MR, Razavi SZ, Langevin R, et al. A virtual conversational agent for teens with autism spectrum disorder: Experimental results and design lessons. In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* 2020: 1-8.
15. Milne M, Raghavendra P, Leibbrandt R, et al. Personalisation and automation in a virtual conversation skills tutor for children with autism. *Journal on Multimodal User Interfaces* 2018; 12: 257-69.
16. Tanaka H, Negoro H, Iwasaka H, et al. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLOS one* 2017; 12(8): e0182151.
17. Bekele E, Wade J, Bian D, et al. Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with Autism spectrum disorders (ASD). *IEEE virtual reality (VR)* 2016: 121-30.
18. Razavi SZ, Ali MR, Smith TH, et al. The LISSA virtual human and ASD teens: An overview of initial experiments. In: International Conference on Intelligent Virtual Agents. Springer, Cham. 2016: 460-3.
19. Serret S, Hun S, Iakimova G, et al. Facing the challenge of teaching emotions to individuals with low-and high-functioning autism using a new serious game: a pilot study. *Molecular autism* 2014; 5: 1-17.
20. Mei C, Mason L, Quarles J. How 3D virtual humans built by adolescents with ASD affect their 3D interactions. In: Proceedings of the 17th international ACM SIGACCESS conference on Computers & accessibility, 2015: 155-62.