Avatar-Based Metaverse Interactions:
A Taxonomy, Scenarios and Enabling Technologies

Hyoseok Yoon[1], Youngho Lee[2], Choonsung Shin[3*]

Abstract

In virtual worlds like online games and metaverse platforms, an avatar represents a user in several graphical representations. In this paper, we offer a taxonomy for avatar-based interactions in the metaverse, highlighting scenarios that are relevant for various features of the used avatars. By using avatar types, flexibility, fidelity, and interaction as our chosen criteria, we try to include recent enabling technologies and existing avatar-based interactions in the taxonomy that is being provided. We explore how current metaverse research, platforms, and interaction technologies might be positioned in various circumstances to demonstrate the viability and usefulness of our approach.

Key Words: Avatar, Metaverse, Taxonomy, Virtual Reality.

I. INTRODUCTION

Users are connecting and staying online for a lot longer now due to recent advancements in computing and the COVID-19 pandemic. A paradigm shift in computing, such as virtual reality (VR), extended reality (XR), and the metaverse, is significant in this regard since it allows users to gradually integrate their real lives into the interconnected virtual worlds known as the metaverse [1-2]. The metaverse roadmap identified four key technical components of the metaverse as virtual worlds, mirror worlds, augmented reality, and lifelogging [2]. In modern metaverse platforms, a user is represented by a graphically produced avatar with a wide range of look and interaction options [3-4]. Researchers and consumers might wonder which approach is viable and effective in different contexts. Indeed, it would be beneficial to present a well-articulated distinctions in the form of a taxonomy to comprehend various research topics and techniques [5]. In this research, to serve many stakeholders in the metaverse by preparing a concrete reference, we propose a taxonomy of avatar-based interaction (ABI) in the metaverse to highlight elements of our taxonomy and to exhibit diverse qualities.

II. RELATED WORK

Avatars have been used to graphically portray users from the real world who are aesthetically dissimilar or similar in appearance since the beginning of metaverse research [6]. Avatar generation, avatar interaction, and the effects of employing avatars have all been the subject of recent studies.

2.1. Avatar Generation

Avatars can already be created from a single image by reconstructing a high-quality, textured 3D face with neutral expressions and normalized lighting [7]. Likewise, Bao et al. have shown off a fully automatic system that uses a consumer RGB-D selfie camera to create high-fidelity, lifelike 3D digital human heads [8]. The avatar's appearance has an impact on training [9]. Such lifelike avatars displayed higher co-presence, interpersonal trust, and increased attention concentration through facial expressions [10].

2.2. Avatar Interaction

A VR communication system was created by Aseeri et al. [11] using an avatar that imitates user gestures, facial expressions, and speech. Additionally, a full-body moving avatar demonstrated the highest co-presence and behavioral dependency [12]. For the augmented reality system, Wang et al. investigated various full-body avatar design types (i.e., body alone, hands and arm, and hands only) [13]. According to Ma and Pan, the cartoon-like avatar made it simpler for participants to manage their facial expressions [14].

2.3. Effects of Employing Avatars

The impact of an avatar's body part visibility on the gameplay and performance of VR games was investigated by Lugrin et al. [15]. A body "avatarization continuum" [16] was presented by Genay et al. to illustrate various levels. The Proteus effect, which describes how people's views and behavior mimic those of their avatars, makes this type of avatar research crucial [17]. Electroencephalography (EEG) was used by Gonzalez-Franco et al. to track participants' brain activity as they interacted with look-alike avatars [18]. While others researched advanced subjects on avatar personalization and visual quality effects, Kegel et al. reported avatar facial expressions that cause diverse brain reactions [19].

III. AVATAR-BASED INTERACTION IN METAVERSE

Our proposed taxonomy for ABI in the metaverse is shown in Fig. 1. Four first-level criteria, including avatar types, flexibility, fidelity, and interaction, make up the taxonomy. To make the taxonomy concise and understandable, there are typically four to five second-level criteria. These characteristics are briefly discussed in this section.

3.1. Avatar Types

The various ways an avatar is graphically generated and displayed on a display are referred to as avatar types [6,8-9]. Users interpret various avatar types based on their visual characteristics.

- 2D Avatars: The avatar is rendered on a two-dimensional (2D) display using a 2D coordinate system. Examples include cartoonish and pixel-based characters.

- 3D Avatars: The avatar is either rendered using a three-dimensional (3D) coordinate system or drawn in a way to visually make the avatar appear in 3D.

- VR Avatars: This type of avatar is associated with the specified 2D or 3D virtual world and constrained by the rendering properties of its virtual world.

- Digital Human: This type of avatar attempts to represent a real user, historical figures, and fictional human-like agents realistically at the highest resolution and quality.

3.2. Flexibility

The various metaverse applications' platform, software, and hardware requirements all affect the current ABI. Additionally, users can create and customize avatars [20]. Flexibility describes both dependency and customization of ABI.

- Hardware Dependency: ABI may use a special hardware (i.e., HMD, smartglasses, smartphone, PC) to perform as intended.

- Software Dependency: ABI may use special software (i.e., Android, macos, Windows, Unity, Unreal Engine) to perform as intended.

- Platform Dependency: ABI is categorized as only usable in a proprietary platform or available across different metaverse platforms (i.e., cross-platforms).

- Customization: ABI is customizable by users. For example, users can change the look and feel of avatars through different items (i.e., clothes, accessories, forms).

3.3. Fidelity

Avatar can be highly detailed or simply abstracted. Fidelity refers to this characteristic of visual properties [21].
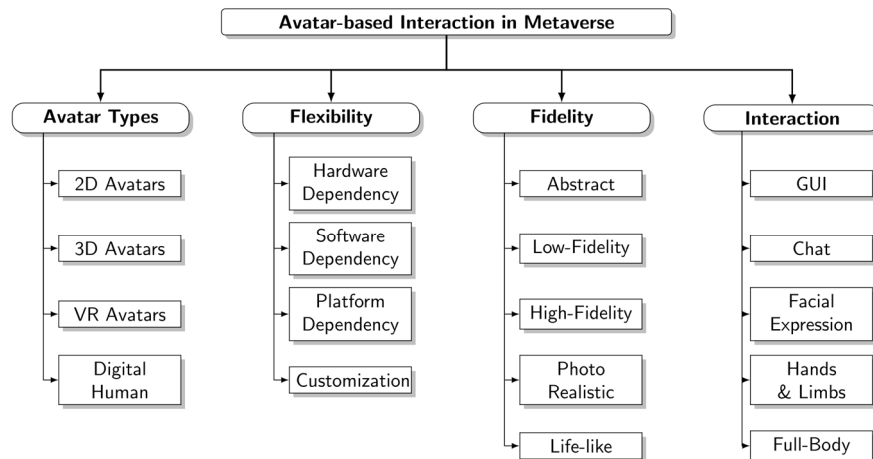

Fig. 1. A taxonomy of avatar-based interaction in metaverse.

- Abstract: The simplest abstraction is used to represent avatars. For example, an avatar may be represented as a dot or with a person icon.

- Low-Fidelity: The avatar is low-fidelity in design if there is a small number of features to distinguish one's avatar from other avatars. In other words, the avatar is limited in its expressivity.

- High-Fidelity: The avatar is high-fidelity in design if there is a greater number of features to distinguish one's avatar from other avatars. In other words, the avatar is highly expressive.

- Photo Realistic: The avatar is photo-realistic if the realistic depiction is used to create an avatar that is self-identifiable.

- Life-like: The avatar is considered life-like if it can behaviorally act like a real person (i.e., walk, run, jump, dance, smile, cry).

3.4. Interaction

Interaction defines the ways an avatar or a user behind the avatar operates an avatar to communicate in the metaverse [25-26].

- GUI: Graphical user interface is a typical interface we use with a WIMP (windows, icons, menus, and pointers) metaphor. For example, an avatar can be controlled using a mouse and a keyboard.

- Chat: An avatar can communicate with text chat or voice/video chat with other avatars. For example, a text bubble or altered voice may be used.

- Facial Expression: An avatar can make various facial expressions such as smile, frown, eyebrow, and lip movements.

- Hands & Limbs: When interacting, users see avatars making gestures and postures using their hands and limbs.

- Full-Body: Avatars can make full-body movements where many body parts are used to make active physical motions.

IV. APPLYING THE TAXONOMY

As shown in Table 1, our proposed taxonomy can be used to describe various metaverse applications [1] and business cases [4]. We briefly discuss 7 scenarios categorized into 3 groups in this section.

4.1. Scenarios

A VR avatar that can be modified with a purchased fashion item may be used in a commerce scenario. A VR avatar that can be modified with a purchased fashion item may be

Table 1. Taxonomy applied to scenarios.

Scenario	Taxonomy criteria*			
	Avtar Types	Flexibility	Fidelity	Interaction
Commerce	3D, VR, DH	CU	HF, LL	GUI, FB
Navigation	2D	H/SD	AB, LF	GUI
Personal	3D, VR, DH	PD, CU	HF, PR	FE, H
Classroom	2D, 3D, VR	H/S/PD	LF	CH, GUI
Workspace	3D, VR, DH	H/S/PD	HF, PR	CH, FE, FB
Social event	VR, DH	CU	LF, LL	FB
Industrial	3D, VR	PD	AB, LF	GUI, HL

*Abbreviation. (DH) Digital Human, (CU) Customization, (HD) Hardware Dependency, (SD) Software Dependency, (PD) Platform Dependency, (HF) High-Fidelity, (LL) Life-Like, (AB) Abstract, (PR) Photo Realistic, (FB) Full-Body, (HL) Hands & Limbs, (CH) Chat, (FE) Facial Expression.

used in a commerce scenario. High-fidelity and full-body interaction must be offered to support these functionalities. We could implement an abstract or low-fidelity GUI-controlled avatar for navigation scenarios. Similar to consumer use cases, industrial use cases do not require workers' avatars to be in-depth because the application's main focus is on the industrial apparatus depicted as a digital twin. Fig. 2 depicts situations for business, navigation, and industry.

Personal and social scenarios are shown in Fig. 3. High-fidelity, digital human, photorealistic, self-identifiable avatars with expressive faces should be offered in personal contexts. The crowds for a social event like concerts can be depicted modestly, but the artists with detailed renderings.

Formal settings located in a workplace and school are shown in Fig. 4. So that multiple users with various hardware, software, and platform requirements can engage in these applications, hardware, software, and platform dependencies should be carefully designed and managed. Additionally,

Fig. 2. Commerce, navigation, and industrial scenarios.

Fig. 3. Personal and social scenarios.

Fig. 4. Formal scenarios (classroom and workplace).

approaches that require reduced computing resources (like as conversation, a graphical user interface, and low-fidelity) can be adopted when an avatar group grows in size. A group of five employees would benefit more from a 3D avatar type, but a classroom of 100 avatars could be implemented with a 2D avatar type.

4.2. Current Metaverse Platforms

Our taxonomy can be used to describe current metaverse platforms and services. Table 2 shows four metaverse platforms (Gather, ifland, Roblox, ZEPETO) characterized by our taxonomy criteria. Currently, all of the metaverse platforms we examined are compatible with running on mobile devices. Due to this hardware dependency, interaction methods are constrained to GUI and chat (text, voice, and video). Another interesting observation is that all of these platforms encourage avatar customization.

4.3. Enabling Technologies

Recent technical developments would be advantageous for the seven scenarios that were covered. Platforms for the metaverse will leverage low-latency broadband networks made possible by 5G and 6G to provide avatars. When connecting the virtual and physical worlds using extended reality and block chains with higher levels of security and protection, rendering that is abstract or highly detailed will need a lot more data to be maintained and transmitted through the cloud. Distributed platforms will improve interoperability and reduce dependency problems in platforms, devices, and software. In addition, media and spatial objects that recreate the portrayal of the real world will be combined with user-representing avatars in the form of digital humans aided by artificial intelligence (AI). To achieve this, Internet-of-Things (IoT), collaborative robots, and digital twins will all contribute to a fully human-in-the-loop metaverse. Avatars will interact with such augmented spatial objects and spaces in the metaverse through real-time user interface (UI) and user experience (UX).

V. CONCLUSION

In this paper, we proposed a taxonomy for avatar-based metaverse interaction. We defined each criterion and

Table 2. Taxonomy applied to current metaverse platforms.

Avtar types	Flexibility	Fidelity	Interaction
	Taxonomy criteria*		
Gather	HD (PC, mobile) CU	LF	GUI (Emote), CH (text, voice, video)
ifland	HD (mobile) CU	LL, HF	GUI (emotions), CH (text, voice)
Roblox	HD (PC, mobile, VR HMD) CU	AB, LF	GUI, CH (text, voice)
ZEPETO	HD (PC, mobile) CU	LL, HF	GUI, CH (text, voice)

*Abbreviation. (CU) Customization, (HD) Hardware dependency, (HF) High-fidelity, (LL) Life-like, (AB) Abstract, (CH) Chat.

applied it to several use cases using the lenses of avatar types, flexibility, fidelity, and interaction. We anticipate that by defining scenarios in concrete terms and laying up the required technologies, our taxonomy can aid researchers working in the field of avatar-based interaction. Researchers can, for instance, investigate the various avatar types and interaction modalities listed in our taxonomy to appropriately situate their services in various application settings.

ACKNOWLEDGMENT

REFERENCES

[1] S. M. Park and Y. G. Kim, "A metaverse: Taxonomy, components, applications, and open challenges," *IEEE Access*, vol. 10, pp. 4209-4251, 2022.

[2] J. M. Smart, J. Cascio, and J. Paffendorf, Metaverse Roadmap Overview, 2007. http://www.metaverseroad map. org/overview/index.html.

[3] Y. Zhao, J. Jiang, Y. Chen, R. Liu, Y. Yang, and X. Xue, et al., "Metaverse: Perspectives from graphics, interactions and visualization," *Visual Informatics*, vol. 6, no. 1, pp. 56-67, 2022.

[4] P. Faraboschi, E. Frachtenberg, P. Laplante, D. Milojicic, and R. Saracco, "Virtual worlds (metaverse): From skepticism, to fear, to immersive opportunities," *Computer*, vol. 55, no. 10, pp. 100-106, 2022.

[5] J. J. Ruscella and M. F. Obeid, "A taxonomy for immersive experience design," in *7th International Conference of the Immersive Learning Research Network (iLRN 2021)*, 2021, pp. 1-5.

[6] E. Schlemmer, T. Daiana, and O. Cristoffer, The Metaverse: Telepresence in 3D Avatar-Driven Digital-Virtual Worlds, *@tic. revista d'innovació educativa*, n.2, 2009. http://ojs.uv.es/index.php/attic/article/view/98/88.

[7] H. Luo, K. Nagano, H. W. Kung, Q. Xu, Z. Wang, and L. Wei, et al., "Normalized avatar synthesis using stylegan and perceptual refinement," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11662-11672.

[8] L. Bao, X. Lin, Y. Chen, H. Zhang, S. Wang, and X. Zhe, et al., "High-fidelity 3D digital human head creation from RGB-D selfies," *ACM Transactions on Graphics*, vol. 41, no. 1, p. 3, 2022.

[9] I. Hudson and J. Hurter, "Avatar types matter: Review of avatar literature for performance purposes," in *Virtual, Augmented and Mixed Reality (VAMR 2016)*, 2016, pp. 14-21.

[10] S. Aseeri and V. Interrante, "The influence of avatar representation on interpersonal communication in virtual social environments, " in *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, 2021,vol. 27, no. 5, pp. 2608-2617.

[11] S. Aseeri, S. Marin, R. N. Landers, V. Interrante, and E. S. Rosenberg, "Embodied realistic avatar system with body motions and facial expressions for communication in virtual reality applications," in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2020)*, 2020, pp. 580-581.

[12] P. Heidicker, E. Langbehn, and F. Steinicke, "Influence of avatar appearance on presence in social VR," *IEEE Symposium on 3D User Interfaces (3DUI 2017)*, pp. 233-234, 2017.

[13] T. Y. Wang, Y. Sato, M. Otsuki, H. Kuzuoka, and Y. Suzuki, "Effect of full body avatar in augmented reality remote collaboration," in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2019)*, 2019, pp. 1221-1222.

[14] F. Ma and X. Pan, "Visual fidelity effects on expressive self-avatar in virtual reality: First impressions matter," *IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 57-65, 2022.

[15] J. L. Lugrin, M. Ertl, P. Krop, R. Klüpfel, S. Stierstorfer, and B. Weisz, et al., "Any "Body" there? avatar visibility effects in a virtual reality game," in *IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2018)*, 2018, pp. 17-24.

[16] A. Genay, A. Lécuyer, and M. Hachet, "Being an avatar "for real": A survey on virtual embodiment in augmented reality," in *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, 2022, vol. 28, no. 12, pp. 5071-5090.

[17] R. Ratan, D. Beyea, B. J. Li, and L. Graciano, "Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the proteus effect," *Media Psychology*, vol. 23, no. 5, pp. 651-675, 2020.

[18] M. Gonzalez-Franco, A. I. Bellido, K. J. Blom, M. Slater, and A. Rodriguez-Fornells, "The neurological traces of look-alike avatars," *Frontiers in Human Neuroscience*, vol. 10, 2016.

[19] L. C. Kegel, P. Brugger, S. Frühholz, T. Grunwald, P. Hilfiker, and O. Kohnen, et al., "Dynamic human and avatar facial expressions elicit differential brain responses," *Social Cognitive and Affective Neuroscience*, vol. 15, no. 3, pp. 68-73, 2020.

[20] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik, "The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, 2018, pp. 1643-1652.

[21] F. Ma and X. Pan, "Visual fidelity effects on expressive self-avatar in virtual reality: First impressions matter," in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2022)*, 2022, pp. 57-65.

[22] K. Takeuchi, Y. Yamazaki, and K. Yoshifuji, "Avatar Work: Telework for disabled people unable to go outside by using avatar robots," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI 2020)*, 2020, pp. 53-60.

[23] F. Miao, I. V. Kozlenkova, H. Wang, T. Xie, and R. W. Palmatier, "An emerging theory of avatar marketing," *Journal of Marketing*, vol. 86, no. 1, pp. 67-90, 2022.

[24] Y. T. Lin, H. S. Doong, and A. B. Eisingerich, "Avatar design of virtual salespeople: Mitigation of recommendation conflicts," *Journal of Service Research*, vol. 24, no. 1, pp. 141-159, 2021.

[25] D. Dewez, L. Hoyet, A. Lécuyer, and F. A. Sanz, "Towards "Avatar-friendly" 3D manipulation techniques:

Bridging the gap between sense of embodiment and interaction in virtual reality," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, p. 264.

[26] N. Ogawa, T. Narumi, and M. Hirose, "Effect of avatar appearance on detection thresholds for remapped hand movements," in *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, 2021, vol. 27, no. 7, pp. 3182-3197.

AUTHORS

Hyoseok Yoon received his B.S. degree in Computer Science from Soongsil University in 2005. He received his M.S. and Ph.D. degrees in Information and Communication (Computer Science and Engineering) from the Gwangju Institute of Science and Technology (GIST), in 2007 and 2012, respectively. He was a researcher at the GIST Culture Technology Institute from 2012 to 2013 and was a research associate at the Korea Advanced Institute of Science and Technology, Culture Technology Research Institute in 2014. He was a senior researcher at Korea Electronics Technology Institute from 2014 to 2019. In September 2019, he joined the Division of Computer Engineering, Hanshin University where he is currently an assistant professor. His research interests include ubiquitous computing (context-awareness, wearable computing) and Human-Computer Interaction (mobile and wearable UI/UX, MR/AR/VR interaction).

Youngho Lee received his B.S. degree in Dept. of Mathematics from KAIST and M.S. degree in the Dept. of Information and communication from GIST. He received his Ph.D. from GIST. Since Sept. 2009, he has been with the Mokpo National University (MNU), where he is a Professor in the Department of Computer Engineering. His research interests include Context-aware computing, HCI, virtual/augmented reality, culture technology, etc.

Choonsung Shin received his B.S. degree in Computer Science from Soongsil University in 2004. He received his M.S. and Ph.D. degrees in Information and Communication (Computer Science and Engineering) from the Gwangju Institute of Science and Technology, in 2006 and 2010, respectively. He was a Postdoctoral Fellow at the HCI Institute of Carnage Mellon University from 2010 to 2012. He was a principal researcher at Korea Electronics Technology Institute from 2013 to 2019 and a CT R&D Program Director of Ministry of Culture, Sports and Tourism from 2018 to 2019. In September 2019, he joined the Graduate School of Culture, Chonnam National University where he is currently an associate professor. His research interests include culture technology & contents, VR/AR and Human-Computer Interaction.