

Towards An Accessible Metaverse: User Requirements

Anna Matamala¹ and Estel·la Oncins¹

¹ *Universitat Autònoma de Barcelona, Edifici K-1002, 08193 Bellaterra, Barcelona, Spain.*

Abstract

The metaverse has the potential to extend the physical world allowing users to seamlessly communicate and interact in a new virtual ecosystem. Yet, it is paramount to ensure that these immersive experiences are accessible to all users regardless of their needs. This paper presents some key aspects to be considered when developing a metaverse for all, reporting on the information gathered to develop two technical specifications approved at ITU Focus Group on metaverse.

Keywords

metaverse, accessibility, users, translation

1. An ecosystem of virtual worlds offering immersive experiences

The metaverse is, according to the definition adopted in December 2023 by the International Telecommunications Union Focus Group on Metaverse, “An integrative ecosystem of virtual worlds offering immersive experiences to users, that modify pre-existing and create new value from economic, environmental, social and cultural perspectives”. Many technologies converge in the metaverse, which includes many virtual spaces with virtual content and virtual people who adopt the form of an avatar. The metaverse is expected to offer experiences in a broad range of fields such as education, healthcare, culture, or shopping, to name a few. The metaverse is in the process of being defined and built [1], so it is a unique opportunity to adopt a born-accessible approach as promoted by Orero [2] and create it in such a way that anyone, regardless of capabilities, can access it [3,4]. This paper aims to highlight some key aspects that could be considered when developing a metaverse for all, reporting on the information gathered to develop two technical specifications approved at ITU: Technical Specification ITU FGMV-04 Requirements of accessible products and services in the metaverse: Part I – System perspective [5] and ITU FGMV-05 Requirements of accessible products and services in the metaverse: Part II – User perspective [6]. The paper adopts a user-centric perspective, describing the actions users may want to perform in the metaverse and what accessibility requirements should be considered (Section 2). Section 3 describes some accessibility services that could be integrated in the metaverse, pointing at some existing research. The article finishes with some conclusions and future work in this field.

2. User actions in the metaverse

There are four main actions that users are expected to perform in the metaverse [6]: a) accessing the metaverse; b) creating an avatar identity in the metaverse; c) navigating in the metaverse, and d) interacting in the metaverse.

¹Joint Proceedings of the ACM IUI Workshops 2024, March 18-21, 2024, Greenville, South Carolina, USA

✉ anna.matamala@uab.cat (A. Matamala); estella.oncins@uab.cat (E. Oncins).

🆔 0000-00002-1607-9011 (A. Matamala); 0000-0002-0291-3036 (E. Oncins).



© 2024 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

2.1. Accessing the metaverse

Virtual worlds are generally accessed through head-mounted displays, but these devices may not cater for the needs of certain users. This is why alternative devices should be considered so that users can choose the one more suitable to their needs or preferences. These devices include hand-based input devices, non-hand-based input devices and motion input devices as explained by Park and Kim [7]. Users with disabilities may use their own assistive technologies to access digital content, hence interoperability between hardware components to access the metaverse and these technologies should be guaranteed in an accessible metaverse. One key aspect when accessing the metaverse is the authentication process, in which specific software may be used. Again, it is paramount that users have alternative options to authenticate themselves, via spoken or written text or via haptics.

2.2. Creating an avatar identity

An avatar is a medium that projects one's identity within virtual spaces [1]. These avatar representations range from self-representations to totally new representations. In other words, users may want to have a faithful depiction of themselves, or they may want to be represented by someone totally different, even a non-human avatar. A metaverse for all should give users a choice of self-expression, allowing them to incorporate a wheelchair or a blind cane in their avatar, should they wish to do so.

2.3. Navigating

Once users have been able to access the metaverse, they are expected to navigate through it perceiving the content that is available. The range of user needs may be varied: a user may need to navigate with a haptic controller, whereas another user may need to navigate relying only on visual input, without access to the audio elements, to put two examples. Software component used in building the virtual worlds should consider accessibility features and interoperability with assistive devices. Some of the features already considered for web-based digital content may also be applicable in this context, such as visual contrast or text magnification, for instance.

2.4. Interacting

Humans are inherently social, so interaction in the metaverse will be central. Similarly to the physical world, interaction can take place by different means: through oral and written language but also through non-verbal communication (i.e., facial expressions, gestures, among other features). This interaction is a bidirectional process in which users can give input and receive responses. To cater for the needs of all users, this interaction cannot stay at only one level but must provide alternative options. For example, a user may want to provide input through spoken words, through keyboards, through gestures or through eye-tracking. Similarly, a user may prefer to receive oral or written responses, depending on their sensory capabilities.

3. Accessibility and translation services towards an accessible metaverse

Accessibility and translation services will play a fundamental role in offering an accessible metaverse, similarly to the key role they already play in the physical and digital world. Next, we present some key services [6] together with some existing research that points at how these access services could be integrated in immersive environments. We would expect the virtual worlds, services, and products to clearly identify the accessibility and translation services available and allow users to customize their choice in an easy way.

3.1. Subtitling

Subtitling can be in the same language (intralinguistic) or in a different language (interlinguistic) and offer a written alternative to the spoken words, hence benefitting those who cannot access the audio for various reasons or those who do not understand the source language. When specifically addressed to persons who cannot access the audio, subtitles also transfer non-speech information such as character or language identification, paralinguistic elements, music, and sounds, among other features. Subtitles follow spatio-temporal constraints to allow users to read the text while enjoying the visuals. Subtitles present different features which users should be able to customize, such as font size, font type, font colour, contrast, or alignment.

Research on subtitling in immersive media started with the BBC study by Brown et al. [8] on four solutions for rendering subtitles and continued with Roth et al. [9] investigations. These studies were the basis for the ImAc project studies [10, 11, 12], which focused on two key aspects: where to place the subtitles in the virtual environment (subtitle positioning), and how to guide the viewers to the speaker in the virtual world (guiding mechanism). When the speaker is not in the field of view of a user who cannot access the audio, a mechanism is needed to guide them. The project tested an arrow and a radar, as guiding mechanisms, and also tested subtitles which were always visible to the user, subtitles attached to the speaker, and subtitles placed in a fixed position every 120°. The always visible subtitles and the arrow were the preferred solutions. More recently, Brescia-Zapata et al. [13] have explored the issues further using eye-tracking to compare always visible subtitles—head-locked in their terminology—versus fixed subtitles and exploring the usefulness of coloured subtitles in three different countries. Although results on subtitle colour are not conclusive, investigations seem to indicate that always visible subtitles are currently the best option to integrate subtitles in virtual environments.

3.2. Transcripts

Transcripts provide a written verbatim alternative to spoken words and may also include some written alternatives to non-speech audio information. For example, one could imagine an oral lecture in an educational metaverse which is transcribed and aligned with the lecturer presentation, which would not only be helpful to students who cannot access the audio but also to anyone wishing to go back to the content. Transcripts are presented in the same language as the audio and, contrary to subtitles, they do not have specific spatio-temporal constraints. Some transcripts highlight certain words as they are spoken, which benefits users with reading difficulties. As in subtitling, one would expect users to be able to customize the features of the transcripts in the virtual world, choosing various features such as font type, colour or contrast. Other questions to be explored might be text position and display, especially in the case of interactive transcripts.

3.3. Audio description

Audio description translates the visual into spoken words [14]. It provides a description of visual elements and some sound elements which may be difficult to understand without access to the visuals. This is especially useful for users who cannot access the visuals and can be applicable to both dynamic and static content. Audio description is offered when the source content has silent spaces. Sometimes it can be offered together with an audio introduction, which is an audio text that provides some key relevant information of the content to be enjoyed before accessing it.

Research on audio description in immersive environments is scarce. As part of the ImAc project [12], three AD modes were tested altering the sound treatment: in a Static mode the sound was located on the user's side, as if someone was sitting close to the user; in the Dynamic mode, the sound was placed where the main action being audio described was taking place; in the Classic mode, the ambisonic sound was placed above the user's head. Although the expectations were that the Dynamic mode would contribute to better understanding the story, the participants felt confused [15] and were more interested in the script characteristics. A second test in Spain and

in the UK focused on these aspects, comparing the so-called Classic, Radio, and Extended AD, always with the same sound treatment. Classic AD offered a standard approach, whereas Radio AD featured a more engaging description. Extended AD allowed users to pause the video and listen to an additional description. British users preferred the Radio approach, whereas Spanish users selected the Extended AD as their preferred option. In any case, these tests show that immersive media may open the door to new AD forms and sound treatment, although unsuccessful in a first attempt, may be an issue worth exploring in certain type of virtual content.

3.4. Audio subtitling

Audio subtitling refers to written subtitles which are converted into spoken words, be it through a text-to-speech system or a human voice. This is especially useful for users who cannot see or cannot read the subtitles and do not understand the source language being subtitled. Audio subtitles can be an independent access service or they can be integrated with audio description. Research on audio subtitles is limited as explained by Matamala [16], but in the field of immersive media is almost non-existent, with some initial tests as part of ImAc [12] focusing only on user preferences regarding the combination of audio description with audio subtitling.

3.5. Interpreting

Interpreting can take place between oral languages (for example, from Catalan into English) or between an oral language and a visual gestural language (for instance, from Catalan into Catalan Sign Language). Interpreting benefit those who do not understand a language or cannot hear it by providing an alternative in another language. Whereas interpreting between oral languages is seen as a translation service, interpreting between oral and sign languages is seen as an access service, but both offer access to a content that would otherwise not be accessible to the user. In an ideal metaverse, users could choose the language of the interpretation together with some other features: volume in the case of oral interpretation or positioning of the sign language interpreter or choice of avatar in the case of sign language interpreting.

Research on interpreting in oral languages in virtual worlds has mainly focused on training and the use of virtual learning environments (VLEs) for collaborative learning in interpreter education. According to results gathered from a pilot test conducted by Braun et al. [17], while VLEs are becoming more accessible due to technological advances and their use has increased rapidly, developing appropriate VLEs to integrating them as a sustainable solution creates further challenges.

As for research on sign language interpreting in immersive media, ImAc pilot testing addressed three main aspects: a) display of signer video, which could be continuous or non-continuous: whereas in the former the signer window was always visible, in the latter it was only present when interpreting was needed; b) presentation of sign language only versus presentation of sign language and subtitles; c) speaker representation, either through an emoji or a textual description [12]. The pilot test took place in Germany with a limited number of users, demonstrating a preference for non-continuous display, simultaneous presentation of sign language and subtitles, and textual description to identify users. Similarly to interpreting, it can be stated that despite of the recent technological advancements for sign language interpreting in immersive environments, there are still unsolved questions to be investigated to increase the effectiveness of communication [18].

3.6. Easy-to-understand language

Easy-to-understand language is an umbrella term to refer to different simplified language varieties (from Easy to Plain Language) that enhance comprehensibility. Easy Language (also called Easy-to-Read) is generally addressed to those who have difficulties reading or understanding language, whereas Plain Language is addressed to all. Although content in the

metaverse may want to play with different language varieties, services addressed to all would be expected to be explained in Plain Language, providing where possible Easy Language alternatives.

Research on easy-to-understand language has generally focused on written texts, with a recent interest in audiovisual content and its relationship with access services [19]. As part of the ImAc project Oncins et al. [20] compared subtitles designed for deaf and hard-of-hearing individuals with simplified subtitles aimed at people with cognitive disabilities in immersive environments. According to results gathered from the pilot test, simplified subtitles were generally preferred as they caused less distraction and permitted greater focus on the primary visual content.

To guarantee access to the metaverse to persons with different cognitive needs, one could suggest offering easy instructions on how to access and navigate the metaverse, together with an easy way to go to a safe place in case they feel overwhelmed, like the quiet rooms that are found in physical spaces.

3.7. Revoicing

Revoicing implies translating a source content and voicing it in another language, be it through dubbing or voice-over. In dubbing the original voices are replaced and there are strict synchrony constraints so that the audience thinks the actors are speaking in the target language [21]. In voice-over, as described by Franco et al. [22], the translated version overlaps with the source version and there are less synchrony constraints.

4. Conclusions

This article has put forward some of the opportunities posed by the metaverse to create a truly accessible virtual world. It has presented some of the existing access and translation services which are available in our physical world and could be transferred in the new ecosystem. Research on the best implementation strategies in the virtual world exists for certain access services but more extensive research is still needed. For instance, the implementation of new access services based on artificial intelligence open the door to a myriad of investigations where users need to be central. The metaverse is for all users to populate, hence user-centric methodologies need to be at the core of future developments.

Acknowledgements

The authors are members of TransMedia Catalonia, a research group funded by the Department of Universities and Research of the Catalan government under the SGR funding scheme (2021SGR00077) and Xarxa AccessCat (2021XARDI00007).

References

- [1] Y. K. Dwivedi, L. Hughes, A. M. Baabdullah, et al. Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int. J. Inf. Manag.* 66, C, 2022. <https://doi-org.are.uab.cat/10.1016/j.ijinfomgt.2022.102542>
- [2] P. Orero, Born accessible: beyond raising awareness, 2020. <https://ddd.uab.cat/record/222130>
- [3] L.-H. Lee, T. Braud, P. Zhou, L. Wang, D. Xu, Z. Lin, P. Hui. All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda, 2021, 10.48550/arXiv.2110.05352
- [4] J. Dudley, L. Yin, V. Garaj. et al. Inclusive Immersion: a review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse. *Virtual Reality*, 27, 2989–3020, 2023. <https://doi-org.are.uab.cat/10.1007/s10055-023-00850-8>

- [5] E. Oncins, C. Eugeni, A. Matamala. ITU Focus Group Technical Specification. Requirements of accessible products and services in the metaverse: Part I – System design perspective. ITU, 2023.
- [6] E. Oncins, C. Eugeni, A. Matamala. (2023) ITU Focus Group Technical Specification Requirements of accessible products and services in the metaverse: Part II – User perspective. ITU, 2023.
- [7] S.-M. Park, Y.-G. Kim. A Metaverse: taxonomy, components, applications, and open challenges. *IEEE Access*, 10, 2022, 4209-4251.
- [8] A. Brown., J. Turner., J. Patterson, A. Schmitz, M. Armstrong, M. Glancy. Exploring subtitle behaviour for 360° video (White Paper whp 330), 2018 <https://www.bbc.co.uk/rd/publications/whitepaper330>
- [9] S. Rothe, K. Tran, H. Hussmann. Dynamic Subtitles in Cinematic Virtual Reality. *Proceedings of the 15th European Interactive TV Conference (ACM TVX 2018)*. ACM, 2018.
- [10] B. Agulló, A. Matamala. Subtitles in virtual reality: guidelines for the integration of subtitles in 360° content. *Íkala*, 25(3), 643-661, 2020.
- [11] B. Agulló, A. Matamala. Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group. *Jostrans. The Journal of Specialised Translation*, 33, 217-235, 2019.
- [12] A. Matamala. Accessibility in 360° videos: methodological aspects and main results of the evaluation activities in the ImAc project. *Sendebarr*, 32, 65-89, 2021.
- [13] M. Brescia-Zapata, K. Krejtz, A. T. Duchowski, C. J. Hughes, P. Orero. Subtitles in VR 360° video. Results from an eye-tracking experiment, *Perspectives*, doi: 10.1080/0907676X.2023.2268122, 2023.
- [14] A. Maszerowska, A. Matamala, P. Orero (Eds) *Audio description. New Perspectives Illustrated*. Benjamins, 2014.
- [15] A. Fidyka, A. Matamala, O. Soler-Vilageliu, B. Arias-Badia. Audio description in 360° content: results from a reception study. *Skase*, 14(1), 14-32, 2021.
- [16] A. Matamala. Audio subtitling. Taylor, Ch.; Perego, E. (Eds) *The Routledge Handbook of Audio Description*. Routledge, 2022.
- [17] S. Braun, E. Davitti, C. Slater. It's like being in bubbles': affordances and challenges of virtual learning environments for collaborative learning in interpreter education. *The Interpreter and Translator Trainer*, 14(3), 2020.
- [18] V. Kasapakis, E. Dzardanova, S. Vosinakis, A. Agelada. Sign language in immersive virtual reality: design, development, and evaluation of a virtual reality learning environment prototype, *Interactive Learning Environments*, 2023. doi: 10.1080/10494820.2023.2277746
- [19] A. Matamala. Easy-to-understand language in audiovisual translation and accessibility: state of the art and future challenges. *X-Linguae*, 15(2), 130-144, 2022. doi: 10.18355/XL.2022.15.02.10.
- [20] E. Oncins, R. Bernabé, M. Montagud, V. Arnáiz Uzquiza. Accessible scenic arts and Virtual Reality. *MonTi: Monografías de Traducción e Interpretación*, 12, 214-241, 2020, <https://raco.cat/index.php/MonTI/article/view/368776>.
- [21] F. Chaume. *Audiovisual translation: dubbing*. St. Jerome Publishing, 2012.
- [22] E. Franco, A. Matamala, P. Orero. *Voice-over Translation: An Overview*. Peter Lang, 2010.