ITUPublications

International Telecommunication Union

Standardization Sector

ITU Focus Group Technical Specification

(03/2024)

ITU Focus Group on metaverse

(FG-MV)

FGMV-26

Requirements for communication between human-avatar languages in the metaverse

Working Group 8: Sustainability, Accessibility & Inclusion

PREPUBLISHED Version

Technical Specification ITU FG-MV-26

Requirements for communication between human-avatar languages in the metaverse

Summary

This Technical Specification provides requirements on how to develop the architecture for communication between humans, digital humans/avatars, and systems in the metaverse. This document considers language modalities, language writing systems, AI language communication technologies, co-linguistic communication, and language prevalence in terms of use. It provides guidance on a wide array of communication workflows for the metaverse. The document also makes recommendations on how communication modalities can be considered in the design of any scenario.

Keywords

Languages, communication, language interaction, AI, translation, minority languages

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1.0 of the ITU Technical Specification on "*Requirements for communication between human-avatar languages in the metaverse*" approved at the 5th meeting of the ITU Focus Group on metaverse (FG-MV), held on 5-8 March in Queretaro, Mexico.

Acknowledgements

This Technical Specification was researched and written by Pilar Orero (Universitat Autònoma de Barcelona, Spain), Rahel Luder (SwissTXT) and Louis Amara (SwissTXT) as a contribution to the ITU Focus Group on metaverse (ITU FG-MV). The development of this document was coordinated by Nevine Tewfik (Egypt) and Pilar Orero (UAB, Spain), as FG-MV Working Group 8 Co-Chairs, and by Yong Jick Lee (Center for Accessible ICT, Rep. of Korea) and Paola Cecchi-Dimeglio (Harvard University) as Co-Chairs of Task Group on accessibility & inclusion.

Special thanks to Nevine Tewfik, Yong Jick Lee, Carrie Chow, Sarah McDonagh, and Wendy Goico Campagna for their helpful reviews and contributions.

Additional information and materials relating to this Technical Specification can be found at: https://www.itu.int/go/fgmv. If you would like to provide any additional information, please contact Cristina Bueti at tsbfgmv@itu.int.

Editor & WG8 Co-Chair:	Pilar Orero UAB Spain	**Tel:** +34 62 275 19 58 **E-mail:** pilar.orero@uab.cat
Editor:	Rahel Luder SWISS TXT AG Switzerland	**Tel:** +41 76 347 50 84 **E-mail:** Rahel.luder@swisstxt.ch
Editor:	Louis Amara SWISS TXT AG Switzerland	**Tel:** +41 79 640 27 91 **E-mail:** Louis.Amara@swisstxt.ch

WG8 Co-Chair:	Nevine Tewfik MCIT Egypt	**E-mail:** ntewfik@mcit.gov.eg
Task Group Co-Chair:	Paola Cecchi-Dimeglio Harvard University	**E-mail:** pcecchidimeglio@law.harvard.edu
Task Group Co-Chair:	Yong Jick Lee Center for Accessible ICT, Korea (Rep. of)	**E-mail:** ylee@caict.re.kr

Table of contents

Page

Technical Specification ITU FGMV-26

Requirements for communication between human-avatar languages in the metaverse

1 Scope

This Technical Specification provides requirements for communication and mapping its architecture in the metaverse. Communication is the basis of any system with human interaction.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Technical Specification. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Technical Specification are, therefore, encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is published regularly. The reference to a document in this Technical Specification does not imply that it has Recommendation status as a stand-alone document.

[FGMV-04] ITU-T FG-MV Technical Specification *"Requirements of accessible products and services in the metaverse: Part I – System Design Perspective"*

[FGMV-05] ITU-T FG-MV Technical Specification *"Requirements of accessible products and services in the metaverse: Part II – User perspective"*

[FGMV-19] ITU-T FG-MV Technical Specification *"Service scenarios and high-level requirements for metaverse cross-platform interoperability"*

[ITU-T F.791] Recommendation ITU-T F.791 (2015), *Accessibility terms and definitions*.

3 Definitions

3.1 Terms defined elsewhere

This Technical Specification uses the following terms defined elsewhere:

3.1.1 Artificial Intelligence (AI) [b-ITU-T M.3080]: Computerized system that uses cognition to understand information and solve problems.

3.1.2 Augmented Reality (AR) [b-ITU-T P.1320]: An environment containing both real and virtual sensory components. The augmented reality continuum runs from virtual content that is clearly overlaid on a real environment (assisted reality) to virtual content that is seamlessly integrated and interacts with a real environment (mixed reality).

3.1.3 Avatar [b-ITU FGMV-D5.1-uriop]: digital representation of a user within a metaverse, and it serves as the primary means of interaction with other users and the metaverse itself. It can be customized to look like the user or take on entirely different forms.

3.1.4 Braille [b-ISO/IEC 17351]: Tactile reading and writing system composed of Braille cells. These are raised dots that can be read with the fingers, especially by people who are blind or who have low vision.

3.1.5 Digital Human [b-ITU-T F.748.15]: A computer application that integrates the technologies of computer graphics, computer vision, intelligent speech and natural language processing. It can be used for digital content generation and human-computer interaction to help improve content production efficiency and user experience.

3.1.6 Diverse users [b-ISO/IEC 71]: Individuals with differing abilities and characteristics or accessibility needs.

3.1.7 Easy-to-understand language [b-ISO/IEC 23859]: Any language variety which enhances comprehensibility. Note 1 to entry: Easy-to-understand language includes plain language, easy language and any intermediate variety. These varieties share many recommendations, but the extent of comprehensibility is different as they address different user needs.

3.1.8 Extended reality: [b-ITU-T P.1320]: An environment containing real or virtual components or a combination thereof, where the variable X serves as a placeholder for any form of new environment (e.g., augmented, assisted, mixed, virtual or diminished reality).

3.1.9 Internet of Things [b-ITU-T Y.4000]: A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.

3.1.10 Mixed reality [b-ISO/IEC 18038]: Merging of real and virtual worlds to generate new environments where physical and synthetic objects co-exist and interact.

3.1.11 Product [b-ISO/IEC 9241-11]: Item that is made or created by a person or machine.

3.1.12 Service [b-ISO/IEC 9241-11]: Means of delivering value for the customer by facilitating results the customer wants to achieve.

3.1.13 System [b-ISO/IEC 9241-11]: Combination of interacting elements organized to achieve one or more stated purposes.

3.1.14 Task [b-ISO/IEC 9241-11]: Set of activities undertaken in order to achieve a specific goal.

3.1.15 User interface [b-ISO/IEC 9241-11]: All components of an interactive system (software or hardware) that provide information and/or controls for the user to accomplish specific tasks with the interactive system.

3.1.16 Virtual reality [b-ISO 9241-394]: Set of artificial conditions created by computer and dedicated electronic devices that simulate visual images and possibly other sensory information of a user's surroundings with which the user is allowed to interact.

3.2 Terms defined in this Technical Specification

3.2.1 Lorm: A tactile alphabet where letters are spelled by tapping or stroking different parts of the hand. Lorm is not universal, each language establishes its conventions.

4 Abbreviations and acronyms

This Technical Specification uses the following abbreviations and acronyms:

AI Artificial Intelligence

CEFR Common European Framework of Reference for Languages

MT Machine Translation

5 Conventions

In this Technical Specification:

The expression "**is required to**" indicates a requirement that must be strictly followed and from which no deviation is permitted if conformance to this Technical Specification is to be claimed.

The expression "**is recommended**" indicates a requirement that is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

The expression "**can optionally"** and "**may**" indicate an optional requirement that is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with this Technical Specification.

6 Introduction

The richness of the world's languages and cultures should be reflected in the metaverse. Languages are part of the world's cultural heritage, with hundreds of languages spoken on every continent and a similar number of sign languages. These languages are an essential part of culture and identity and reflect the history of each community.

Communication in these languages can also take place in different modalities: oral languages, sign languages, texts such as subtitles or bots, and easy to read. The metaverse should be multilingual and multimodal by default to avoid exclusion and promote diversity.

Communication in the metaverse will be beneficial for economic prosperity. In today's globalised world, language skills are essential for economic success. The UN has a comprehensive policy framework for work equity, which is based on the principle of equal treatment for all people, regardless of gender, race, ethnicity, religion, sexual orientation, or disability. Communication in the metaverse, supported by Artificial Intelligence (AI), should be multilingual and multimodal by design, and this has direct implications for the architectural design of the underlying technology.

6.1 Motivation

Communication in the physical world poses barriers to many users. The metaverse offers a unique opportunity to create an accessible virtual world if communication requirements are considered in the architecture design. This requires consideration of the needs of a wide range of users. These include, but are not limited to, users who may experience challenges accessing audio or visual content, reading and understanding written language, understanding oral language, speaking, touching, using fine motion, and moving the body or parts of it.

6.2 Communication considerations

Communication in the metaverse can go in several directions:

- The human accesses the metaverse
- The human creates an avatar language identity (e.g. a deaf person using a) sign language and b) Korean sign language) in the metaverse
- The human communicates with the avatar/digital human
- The avatar communicates with other avatars
- The avatar communicates with the system/digital human
- The system/digital human communicates with the avatar
- The avatar communicates with the human

Many examples of metaverse communication situations can be found in the ITU-T FG-MV Technical Specification "*Service scenarios and high-level requirements for metaverse cross-platform interoperability*" [FGMV-19]

7 Existing communication interactions

There are many ways to communicate:

7.1 Language modality

7.1.1 Oral languages

These are the most common and include most spoken languages, such as English, Spanish, Mandarin Chinese, and French. These languages are produced through spoken sounds and are perceived through the auditory system.

7.1.2 Visual languages

These are produced by handshapes, body postures, and facial expressions, and are perceived through the visual system. Sign languages, such as American Sign Language (ASL) and British Sign Language (BSL), are two examples of visual-gestural languages.

7.1.3 Tactile languages

These languages are produced by touch and are perceived through the somatosensory system. They are typically used by deaf and blind people. Two examples of tactile alphabets used for communication in tactile languages are Braille and Lorm. Sighted people recognize and process letters by their visual characteristics, whereas in Braille and Lorm reading uses the somatosensory system for letter perception.

7.1.4 Textual languages

These languages are produced by written words or symbols and are perceived by the visual system.

7.1.5 Brain Computer Interface (BCI)

BCI is a field of technology that aims to create a direct communication pathway between the brain and external devices, such as computers or robotic limbs. By recording and interpreting electrical signals generated by the brain, BCIs allow individuals to control these devices without the need for physical movement. There are two main types of BCIs:

- EEG-based BCIs: These BCIs measure electrical brain activity using electrodes placed on the scalp. They are the most common type of BCI and are used for a variety of applications, such as controlling cursor movements on a computer screen or operating a robotic arm.
- ECoG-based BCIs: These BCIs measure electrical brain activity using electrodes placed directly on the surface of the brain. They are more invasive than EEG-based BCIs, but they can provide more precise information about brain activity.

7.1.6 Augmentative and Alternative Communication.

Augmentative and Alternative Communication is a broad term that covers a wide range of strategies, techniques, and tools that can be used to augment or replace speech or writing for people who have difficulty communicating. These tools can be used by people of all ages and with a wide range of communication needs.

Non-electronic Augmentative and Alternative Communication strategies:

- Gestures: Gestures are natural and intuitive ways to communicate, and they can be used to express a wide range of ideas. For example, pointing can be used to indicate objects or locations, swiping for navigation, tapping for selection, scrolling for movement, rotation for adjustment, nodding to indicate agreement, and shaking the head can be used to indicate disagreement.
- Picture communication boards: Picture communication boards are a collection of pictures that can be used to represent words or phrases. The user can point to the pictures to communicate their needs or wants.
- Alphabet boards: Alphabet boards are a device that displays the alphabet in alphabetical order. The user can point to the letters to spell words.
- Eye gaze communication systems: Eye gaze communication systems use a computer to track the user's eye movements. The user can select words or phrases by looking at them on the screen.

Electronic Augmentative and Alternative Communication strategies:

- Speech Generating Devices are devices that produce speech. The user can type words or phrases into the device, and the device will speak aloud to them.
- Speech Generating Devices and Voice Output Communication Aids : both are similar , but they are designed for people who have difficulty using their hands. The user can activate the Voice Output Communication Aids by using a switch or a scanning device.
- Touchscreens: Touchscreens allow the user to interact with a device by touching the screen. This can be a useful option for people who have difficulty using their hands for typing or who have visual impairments.
- Touch screens Augmentative and Alternative Communication strategy
- Augmentative and alternative communication software: There is a wide range of this software available, which can be used on computers, tablets, and smartphones. This software can provide a variety of communication tools, such as picture communication boards, symbol sets, and word prediction features.

7.2 Language writing systems

A writing system is a system of conventional graphic symbols that represent the spoken language of a particular country or community [b-Crystal]. Reading in a writing system may be from top to bottom, from left to right, or from right to left. This has direct implications in the metaverse when translating from one writing system to another, for example, from English to Japanese. It is also important to take this into consideration when placing or displaying any written text, such as subtitles/captions. There are five main writing systems [b-Coulmas]:

1. Logographic writing systems use individual symbols to represent words or concepts. For example, the Chinese writing system uses more than 50,000 logographic symbols.
2. Syllabic writing systems use symbols to represent syllables, which are combinations of consonants and vowels. For example, the Japanese writing system uses a combination of logographic characters, called kanji, and syllabic characters, called hiragana and katakana.
3. Alphabetic writing systems use symbols to represent individual sounds, or phonemes. For example, the English writing system uses 26 letters to represent the sounds of the English language.
4. Abugida writing systems are a type of alphabetic writing system in which each symbol represents a consonant and an inherent vowel. For example, the Amharic writing system, used in Ethiopia.
5. Abjad writing systems are a type of alphabetic writing system in which each symbol represents a consonant. Vowels are usually not written explicitly but are indicated by diacritics or by contextual clues. For example, the Arabic writing system is an abjad.

In addition to these five main types of writing systems, there are also a number of other writing systems that do not fit neatly into any one category. For example, the Cherokee writing system, developed by Sequoyah in the early 19th century, is a unique system that combines logograms, syllabograms, and alphabetic symbols.

7.3 AI language technologies for communication

Artificial intelligence (AI) has made significant advances in the field of natural language processing (NLP), enabling machines to understand and generate human language. These AI language technologies are continuously evolving, enabling machines to better understand and interact with human language. As these technologies mature, they will play an increasingly significant role in shaping the future of communication and education. Here are some of the main AI language technologies that are shaping the future of communication and interaction:

- **Machine Translation systems:** use AI algorithms to translate text from one language to another. They have become increasingly sophisticated, enabling accurate translations for a wide range of languages.
- **Neural Machine Translation systems:** use artificial neural networks to translate text. They have surpassed traditional statistical MT methods and are now the dominant approach in machine translation.
- **Speech Recognition systems:** enable machines to understand spoken language. They are used in voice assistants, dictation software, and other applications that rely on voice input. Examples of these systems include Apple's Siri or Amazon's Alexa.
- **Natural Language Generation systems:** systems can generate text that resembles human language. They are used in chatbots, virtual assistants, and text summarization tools.
- **Question Answering systems:** can answer questions asked in natural language. They are used in search engines, virtual assistants, and educational applications.
- **Sentiment Analysis systems:** can analyze text to determine its emotional tone. They are used in customer service, social media analysis, and marketing campaigns.
- **Text Summarization systems:** can automatically generate summaries of text documents. They are used in news articles, research papers, and legal documents.
- **Text Classification systems:** can categorize text documents based on their content or topic. They are used in email filtering, spam detection, and content recommendation.
- **Text Embedding systems:** represent text as numerical vectors that capture semantic meaning. They are used in NLP tasks such as similarity detection, sentiment analysis, and MT.
- **Natural Language Understanding:** can analyze and interpret natural language input, enabling machines to understand the meaning and intent of human communication. They are the foundation for many AI language applications.

7.4 Co-linguistic communication

Human communication can combine different language modalities, and linguistic backgrounds, also known as code-switching. For example, in Switzerland, one person may speak in French, and the other person may respond in Italian synchronically with Italian Sign Language. It is a complex process involving a number of factors, including language skills, cultural awareness, and non-verbal communication. It is a common practice in bilingual or multilingual countries, which are the majority in the world.

7.5 Language prevalence in terms of use

The size of a language has a significant impact on the development and use of technology. For example, the development of Large Language Models (LLMs) is compromised by the availability of a massive amount of training data in a given language. Languages with small populations of speakers will produce smaller language models, for example Korean has around 80 million speakers compared to Norwegian with close to 4 million. In addition, the complexity of larger languages can make it more difficult to develop language-specific APIs and tools. For example, it may be more difficult to develop a MT system for a language with complex grammar rules or a large number of homonyms.

8 Common requirements for communication between human-avatar languages in the metaverse

When designing communication interactions between human-avatar languages in the metaverse, the following requirements need to be considered:

- It is required to design any metaverse application with communication choices to be included in products and services.

  Note 1 – for details, see [FGMV-05].

- It is required to design any implementation for the metaverse to allow for language modality and its accessibility services.

  Note 2 – for details, see [FGMV-04].

- It is required to design an architecture for the metaverse with interoperability with the user's preferred/needed options for activation, deactivation and/or pausing of the accessibility services, languages, and language modalities.

  Note 3 – for details, see [b-ISO/IEC 71].

- It is recommended to allow for multiple choices for services, languages, and modalities simultaneously.

  Note 4 – for details, see [b-ISO/IEC 71].

Bibliography

[b-ITU-R BT.2207-6] ITU-R BT.2207-6 (2022), *Accessibility to broadcasting services for persons with disabilities*.

[b-ITU-R BT.2447-2] ITU-R BT.2447-2 (2021), *Artificial intelligence systems for programme production and exchange*.

[b-ITU-R BT.2420-5] ITU-R BT.2420-5 (2022), *Collection of usage scenarios of advanced immersive sensory media systems*.

[b-ITU-T J.301] Recommendation ITU-T J.301 (2014), *Requirements for augmented reality smart television systems*.

[b-ITU-T Y.4000] Recommendation ITU-T Y.4000/Y.2060 (2016), *Overview of the Internet of things*.

[b-ITU-T P.1320] Recommendation ITU-T P.1320 (2022), *Quality of experience assessment of extended reality meetings*.

[b-ITU-T F.748.15] Recommendation ITU-T F.748.15 (2022), *Framework and metrics for digital human application systems*.

[b-ITU-T M.3080] Recommendation ITU-T M.3080 (2021), *Framework of artificial intelligence enhanced telecom operation and management (AITOM)*.

[b-ITU-T P.1320] Recommendation ITU-T P.1320 (2022), *Quality of experience assessment of extended reality meetings*.

[b-ITU FGMV-D5.1-uriop] ITU FGMV-D5.1-uriop (2024), *Service scenarios and high-level requirements for metaverse cross-platform interoperability*.

[b-ISO 17351:2013] ISO 17351:2013, *Packaging — Braille on packaging for medicinal products*.

[b-ISO 9241-394] ISO 9241-394: 2020, *Ergonomics of human-system interaction – Part 394: ergonomic requirements for reducing undesirable biomedical effects of visual induced motion sickness during watching electronic images*.

[b-ISO/IEC 2382] ISO/IEC 2382:2015, *Information technology – Vocabulary*.

[b-ISO/IEC 9241-11] ISO/IEC 9241-11:2018, *Ergonomics of human-system interaction – Part 11: Usability: definitions and concepts*.

[b-ISO/IEC 18038] ISO/IEC 18038: 2020, *Information technology – Computer graphics, image processing and environmental representation – Sensor representation in mixed and augmented reality*.

[b-ISO/IEC 23005-4] ISO/IEC 23005-4:2018, *Information technology — Media context and control — Part 4: Virtual world object characteristics*.

[b-ISO/IEC 23859] ISO/IEC 23859:2023, *Information technology – User interfaces – Requirements and recommendations on making written text easy to read and understand*.

[b-ISO/IEC 71] ISO/IEC 71:2014, *Guide for addressing accessibility in standards*.

[b-Crystal] Crystal, David (2008) *A dictionary of Linguistics and phonetics*. Wily Blackwell.

[b-Coulmas] Coulmas, Florian (1996). *The Blackwell Encyclopedia of Writing Systems*. Oxford: Blackwell Publishers Ltd. ISBN 0-631-21481-X.