**Source:** China Mobile

**Title:** **[iRTCW]Functional Requirements for** **Avatar** **Driven**

**Document for** Agreement

**Agenda item:** 10.5

# Introduction

TR 26.928 identified two use cases and requirements on AR avatar, i.e.AR animated avatar calls and AR avatar multi-party calls (A.18 and A.19 in TR 26.928). The AR glasses user sees an animated avatar of the phone user. Movements of the phone user are used to control the animation of his avatar. These two use cases illustrate the requirements for avatar-driven technology, that is, the driving technology can be separated from 3D video capture, driving an off-the-shelf 3D image instead of capturing it in real time during a call. With the real-time driving , the avatar can behave like a human and have the ability to express with language, facial expressions and body movements.

This document evaluates some requirements to address support for avatar driven for iRTCW, since several OTT services conducted by avatars have realized that movements and expressions are synchronized with reality. Real-time driving technology support various complex interactions in the immersive space. In addition to relying on infrastructure such as computing resource and network, it is also a combination of various technologies such as modeling, rendering, driving, and AI. More over, it relies on real devices to display. The main implementation methods include human driven and AI driven.

# Workflow for Avatar Driven

## 2.1 Human-Driven

Human-Driven refers to capturing the action and facial expression data of live-action people, and then transferring these data into a virtual avatar.

## With the progress of motion capture, posture, expression and other recognition algorithms, ordinary cameras can achieve more accurate motion (for example, iPhone 12 camera can already support simple motion capture).The general workflow is as follows.

1.Avatar generation or selection. Users can design or load avatar and multiple avatar roles can be selected for the user to log on.

2.Model binding. It includes conducting 3D modeling of face and body, selecting key points and mapping the identified key points to the model.

3.Motion capture. It will capture key changes in body shape, expression, eyes, gestures, etc.

4.Driving and rendering.

5. Generate content and interact.

Therefore, the requirements of the client supporting human-driven may be as follows

- Support for driving mode selection.

- At least one RGB camera for input.

- Support for human-driven applications.

- Support expression-based or other model-driven data and interface requirements.

## AI-Driven

For AI-driven, speech, facial expression and specific action of the virtual avatar will be driven in real time mainly through the calculation results of a deep learning model, and the final effect will appear after rendering.The general workflow is as follows.

1.Avatar generation or selection.

2.Model binding.

3.Train various driving models to determine the final effect.

4.Content production. Based on input speech (or speech transformed from input text), predict lip movement, expression and other parameters.

5.Render and generate the final content.

Therefore, the requirements of the client supporting AI-driven include

- Support for driving mode selection.

- One microphone or a keyboard for input.

-Support for AI-driven applications

-Support TTS(Text To Speech) and interface requirements

-Sufficient computing resource to support AI and rendering, ensuring real-time and high-precision user experience.



Figure 1  the process of avatar driven

# Proposal

It is proposed to make the cameras, microphones, sensors, and other features of the iRTC client to support these avatar-type and driving applications.The following next steps should be taken:

-Specify the requirements for avatar-type formats.

-The implementation methods of avatar driven can be taken into account proposed in Section 2.

# References

[1] NVIDIA Omniverse Audio2Face <https://www.nvidia.com/en-us/omniverse/apps/audio2face/>

[2]ARKit <https://developer.apple.com/documentation/arkit/content_anchors/tracking_and_visualizing_faces>

[3]Horizon World metaverse

[https://www.oculus.com/horizon-worlds/?utm\_source=gg&utm\_medium=ps&utm\_campaign=17154134368&utm\_term=horizon%20world&utm\_content=&gclid=CjwKCAjwi8iXBhBeEiwAKbUofcHZg5PB7YylkYC0](https://www.oculus.com/horizon-worlds/?utm_source=gg&utm_medium=ps&utm_campaign=17154134368&utm_term=horizon world&utm_content=&gclid=CjwKCAjwi8iXBhBeEiwAKbUofcHZg5PB7YylkYC0KbkcjxRZk6gKQVf9KvWaUAQvGFsVbra5n5IMSRoCRZ4QAvD_BwE&gclsrc=aw.ds)