**Agenda item:** 10.7

**Source:** Qualcomm Inc.

**Title: [AvCall-MED] 2D Avatar Profile**

**Document for** Discussion andAgreement

# Introduction

In this contribution, we propose a definition for the 2D avatar profile of MPEG ARF, which is defined for usage with 2D avatars.

# Proposed changes

### 5.6.3 2D Avatar Format

An AR-MTSI client that supports 2D avatars shall support the following data formats and requirements for the avatar assets:

* 2D mesh representations that consist of a single planar mesh or quad suitable for texture mapping, conforming to the binary glTF (GLB) format,
* Static image assets for the base avatar representation that conform to still image formats as defined in section 5.5, with support for still image formats as defined in TS 26.114,
* Landmark sets for facial animation, where landmarks shall be defined as 2D coordinates (x, y) on the avatar image plane,
* Texture data components that contain the avatar appearance, which shall conform to still image formats as defined in section 5.5,
* Animation data based on landmark positions, where the landmark animation samples shall use the 2D coordinate format (ala\_is\_3d\_flag set to false) as specified in clause 8.4 of ISO/IEC 23090-39 [11], and
* Sparse tensor data formats for landmark vertex indices as described in the ARF specification [11].

For 2D avatars, the ARF document shall include:

* At least one landmark set that defines facial feature points for animation,
* Support for the landmark animation type in the supportedAnimations list, identified by an appropriate URN,
* A simplified asset structure with a single level of detail containing the 2D mesh and associated texture, and
* LandmarkSet components that reference vertices on the 2D mesh corresponding to key facial features (e.g., eyes, mouth, eyebrows).

The 2D avatar animation should be achieved through landmark-based deformation, where:

* Landmark positions from the animation stream drive the deformation of the 2D mesh,
* The mesh deformation should preserve the natural appearance of the facial features, and
* The number of landmarks should be sufficient to capture essential facial expressions while maintaining computational efficiency.
* Data items of the 2D avatar should signal no compression or protection schemes to ensure broad compatibility.

Alternatively, support for animation using voice-based animation through a pre-trained and fine tuned model for the user is possible. The DNN model is stored as part of the ARF container and declared as proprietary animation framework.

NOTE: The specific landmark configuration and animation URNs for 2D avatars may be defined in operator-specific profiles or through industry fora.

# Proposal

We propose to agree the content of section 2 to the base CR.