**3GPP SA4 RTC SWG AH *S4aR250136r01***

**Paris, FR, 3–5 Sept 2025**

**Agenda item:** 7

**Source:** Nokia

**Title: [AvCall-MED] On animation source streams**

**Document for** Discussion andAgreement

# Introduction

The present document discusses the source of animation data in an avatar call and how to ensure robustness during the session using voice-driven animation.

In a basic case a UE1 with an avatar will transmit an animation stream to the network or a UE2 where the animation stream is used to animate the base avatar of UE1 user. The animation stream is typically created using sensor input from the Party A, which may include camera, microphones (for voice-driven avatars) or other sensors.

When such sensory input is temporarily unavailable (e.g., the user moves out of view of the camera in cases of user video-based animation data generation, leaves the phone on the table while doing something else, etc.), voice can be used as fallback. It is possible for the sender (e.g., a UE or MF) of the animation data to switch to a new source data (e.g., to voice-driven animation) for creating the animation data. However, in some cases the remote UE2 may do the voice animation locally when, for example:

* UE1 and network cannot generate animation data from voice. If voice is transported UE-to-UE, the network may not have access to voice data for generating the animation data.
* There is congestion on the network. Sending voice and animation data based on voice creates redundancy if the remote UE is capable of generating animation data based on voice.

The required signalling for avatar calls can be divided into two categories, excluding the IMS data channel and media configurations that are inherited from MTSI calls:

* The initial SDP exchange required to communicate the capabilities of the UE and the network to establish the parties having an avatar within the call, the format of the avatars, the entities where these avatars will be animated and rendered and the animation streams.
* Management of the avatar scene via scene descriptions and other metadata exchanged during the call.

This contribution proposes required signalling for voice fallback during an avatar call.

# Identification of fallback animation source in DC negotiation

Include a fallback property in the animationFramework component defined in the ADC negotiation [S4aR250142].

**Table 9 — Component: animationFramework**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Type** | **Use** | **Description** |
| offered | Array<string> | C | Framework URNs offered by the sender (for example, urn:3gpp:ar:anim:face.blendshape.v1, urn:3gpp:ar:anim:skeleton.v1, urn:3gpp:ar:anim:landmarks2d.v1,  Urn:3gpp:ar:anim:voice); present in offer. |
| selected | string | C | Selected framework URN; present in answer, commit, and any update that changes the framework. |
| fallback | string | C | Fallback framework URN; present in answer, commit, and any update that changes the framework. |

Fallback is an audio stream that is to be used for animating the avatar in the absence of an avatar animation stream.

When fallback is used, the SDP media description of the fallback audio stream shall include a MID value.

The MID value of the fallback stream shall be sent as part of the suspend/pause message over ADC (defined in clause 3.below).

# Sending notification for animation data disruption

An AR-MTSI client that offers an avatar animation stream should notify the remote client with a message over the data channel with the URN **urn:3gpp:avatar:animation:stopped** when the animation stream becomes temporarily unavailable**.** An AR-MTSI client that receives the message shall switch to audio-driven animation if a source audio stream whose media description includes the SDP attribute avatar\_voice\_fallback was negotiated. Format of the stopped message is shown in Table 1.

Table 1: **Stopped** message for avatar animation

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Type** | **Cardinality** | **Description** |
| **id** | string | 1..1 | A unique identifier of the message in the scope of the data channel session. |
| **type** | string | 1..1 | A URN that identifies the message type: **urn:3gpp:avatar:animation:stopped** |
| **reason** | string | 0..1 | An optional field indicating the cause of disruption in animation stream. Possible values include: "device error", "low-confidence for sensor data", "network issues", etc. |
| **startTime** | number | 0..1 | start time of the suspension of the animation data |
| **endTime** | number | 0..1 | end time of the suspension of the animation data if known |
| **mid** | number | 0..1 | MID value of the audio stream used as fallback when animation stream stopped. |

When the animation stream becomes available again, the AR-MTSI client that offers the avatar animation stream should notify the remote client with a message over the data channel with the URN **urn:3gpp:avatar:animation:resumed**.

Table 1: **Resumed** message for avatar animation

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Type** | **Cardinality** | **Description** |
| **id** | string | 1..1 | A unique identifier of the message in the scope of the data channel session. |
| **type** | string | 1..1 | A URN that identifies the message type: **urn:3gpp:avatar:animation:resumed** |

# Proposal

Section 2 and 3 should be added to appropriate sections in the CR to TS 26.264.