**3GPP TSG-SA WG4 Meeting #130 *S4-241874\_r03***

**Orlando US, 18th Nov – 22th Nov, 2024 *Merge with S4-241948***

**Source: HUAWEI, Samsung Electronics Co., Ltd.**

**Title: [FS\_AVATAR] pCR on IMS architecture mapping**

**Spec: 3GPP TR 26.813 v0.7.0**

**Agenda item: 9.7**

**Document for: Discussion and agreement**

**1. Introduction**

During the previous 129-e meeting discussion, it was asked to suspend the study of IMS mapping for avatar communication until the conclusion from SA2 is available. Since the conclusion of the architecture from SA2 NG\_RTC\_Ph2 has been reached in S2-2411010 as agreed in last SA2#166 meeting, this contribution proposes to update IMS architecture mapping for avatar communication.

According to the comments received during SA4#130, it’s proposed to merge with the content of S4-241948 regarding the architecture mapping part. Also an editor’s note is given to ensure consistent alignment with SA2’s ongoing work.

**2. Reason for Change**

Update IMS architecture mapping for avatar communication to align with the SA2 conclusion.

**3. Proposal**

It is proposed to agree on the following changes to the latest TR 26.813 v0.7.0 (and also update the corresponding clause 6.3.1 in the latest PD document v0.5.0 in sync).

\* \* \* 1st Change \* \* \* \*

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TR 22.856: “Feasibility Study on Localized Mobile Metaverse Services”

[3] ISO/IEC 23090-14:2023, Information Technology – Coded Representation of Immersive Media – Part 14: MPEG Scene Description.

[4] AFLW (Annotated Facial Landmarks in the Wild), accessible here: <https://paperswithcode.com/dataset/aflw>, January 2024.

[5] The LFPW (Labeled Face Parts in the Wild)，accessible here: <https://paperswithcode.com/dataset/lfpw>, July 2023

[6] WFLW (Wider Facial Landmark in the Wild), accessible here: <https://paperswithcode.com/dataset/wflw>， November 2023

[7] WG03N1316, “Procedures and Test Formats for Avatar Representation Formats as part of MPEG-I”, MPEG#147, Sapporo, Japan, July 2024.

[8] Wang, Feng, Hang Zhou, Han Fang, Xiaojuan Dong, Weiming Zhang, Xi Yang and Nenghai Yu. “Deep 3D mesh watermarking with self-adaptive robustness.” Cybersecurity 5 (2021): n. pag.

[9] Zhu, Xingyu, et al. "Rethinking Mesh Watermark: Towards Highly Robust and Adaptable Deep 3D Mesh Watermarking." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. No. 7. 2024.

[10] Narendra, Modigari, et al. "Levenberg–Marquardt deep neural watermarking for 3D mesh using nearest centroid salient point learning." Scientific Reports 14.1 (2024): 6942.

[11] ITU-T F.748.14 (03/2022), "Requirements and evaluation methods of non-interactive 2D real-person digital human application system".

[12] ITU-T F.748.15 (03/2022), "Framework and metrics for digital human application system".

[13] Zhou Y, Han X, Shechtman E, et al. Makelttalk: speaker-aware talking-head animation[J]. ACM Transactions On Graphics (TOG), 2020, 39(6): 1-15.

[14] Zhao J, Zhang H. Thin-plate spline motion model for image animation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 3657-3666.

[15] Liu, Jinglin, et al. "Parallel and High-Fidelity Text-to-Lip Generation." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 2. 2022.

[16] PyTorch, https://pytorch.org/

[17] ONNX, https://onnx.ai/

[18] TensorRT, https://developer.nvidia.com/tensorrt

[19] mindspore, <https://www.mindspore.cn>

[20] https://www.theverge.com/2018/11/8/18074806/ai-news-anchor-china-xinhua-digital-composite

[21] <https://www.tiktok.com/@chriscanbefunny/video/6902602432811961605>

[22] Yilong Liu, et al., “Video-Audio Driven Real-Time Facial Animation”, https://www.cs.sjtu.edu.cn/~shengbin/course/SE/Video-Audio%20Driven%20Real-Time%20Facial%20Animation.pdf

[23] Diana T. Mosa, et al, “A real-time Arabic avatar for deaf–mute community using attention mechanism”, <https://link.springer.com/article/10.1007/s00521-023-08858-6>

[24] 3GPP TR 26.927 v0.7.0: “Study on Artificial Intelligence and Machine Learning in 5G”.

[25] 3GPP TS 23.228: "IP Multimedia Subsystem (IMS); Stage 2".

\* \* \* 2nd Change \* \* \* \*

## 8.6 Mapping to IMS-based Services

### 8.6.1 Architecture Mapping



**Figure 13: Mapping Avatar Functions to IMS DC Architecture**

Annex AC of TS 23.228 [25] describes the IMS DC architecture for avatar communication. As a supplement to it, figure 8.6.1-1 shows the mapping of avatar functions which are defined in Figure 11 of clause 7 to the IMS DC architecture, specifically the possible avatar functions which may be supported by the MF. Note that whilst not shown in Figure 8.6.1-1, the Animation Data Generation, Avatar Animation, and Base Avatar Generation functions may also be part of the UE.

The generation of a Base Avatar by the Base Avatar Generation function may happen in either the UE or the MF, but Base Avatars may also already be available in the Avatar Storage function either in the UE or the Base Avatar Repository (BAR). Base Avatars generated by the UE or the MF may be stored into the UE or the BAR.

Depending on the possible configurations, as shown in Figure 12 in clause 7, a specific avatar workflow is decided through the negotiation between the UE and the network, ultimately deciding on the need for certain avatar functions in each entity.

**BAR (Base Avatar Repository):**

- Avatar Storage: Stores the Base Avatar Representations and their associated Avatar IDs.

NOTE: One or more Base Avatars may be stored for a user, and each Base Avatar is identified with an Avatar ID.

**MF:**

- Base Avatar Generation: the MF may generate base avatar by the user input and stores the base avatar to BAR. For 3D avatars, the base avatar may be a 3D model or an INR model. For 2D avatars, the base avatar is comprised of a DNN model and a base image/video.

- Animation Data Generation: the MF generates animation data using conventional or AI/ML technologies based on the media received from the user.

- Avatar Animation: the MF generates or downloads the base avatar, and animates the base avatar using the received animation data.

- Base Avatar Cache: Base Avatars generated by the UE or the MF may be stored into the UE or the BAR. During an IMS-based avatar communication, the MF may temporarily store relevant Base Avatars in its Base Avatar Cache for provision to participating UEs.

**DC AS:**

- Scene Management: supports the scene description document management. For 2D avatar, the scene description is not needed.

Through such functions, the network may assist the UE with media processing related to the creation of avatar and animation data, as well as the consumption of avatar data, in particular scene management/composition and rendering.

For the support of avatar services based on the IMS DC architecture, media negotiation between the UE and the network should include aspects related to:

* UE capability
* Network capability

The following media interface are used for the IMS-based avatar communication services.

- MDC4: Reference point of Avatar representation downloading between MF and BAR.

**Editor’s note:** this section needs to be revisited after SA2 finalizes their relevant work to ensure the consistency of the avatar functions mapping to the IMS DC architecture.

\* \* \* End of Changes \* \* \* \*