**Source: Beijing Xiaomi Mobile Software Co., Ltd**

**Title:** **Proposal of a** **new Example Usage Scenario**

**Document for: Discussion & Agreement**

**Agenda Item: 7.5**

1. **Summary**

It is agreed that IVAS Usage Scenarios (IVAS-9)[1] provides a collection of example usage scenarios for IVAS[2], the source proposes a different usage scenario using template format proposed in Annex A. Example on Section 2 introduces an example usage scenario for remote class participation.

1. **Immersive and focused remote class participation Example Usage Scenario**

The source suggests the addition of the following table to section 3.2.4 to the IVAS-9 P-doc based on template proposed in Annex A.

|  |
| --- |
| **Usage Scenario Name** |
| Immersive and focused remote class participation |
| **Description** |
| The immersive and focused remote class participation scenario is preferable to have functions of   1. multiple participants connected with various devices. 2. stereo/multiple channel rendering or binaural rendering 3. speech from a 3DoF or 6DoF speaker can be heard by all participants or a particular participant headed by the speaker.   In this usage scenario, 3DoF means the speaker is in a fixed position, but the speaker can change their orientation. 6DoF means the speaker's position and orientation can both be changed. The students can only do activities in a 3DoF style and their positions are predefined initially. These predefined positions in virtual classroom are similar to the seats in a physical classroom, which is initialled by the teacher. The teacher could do activities in a 6DoF style. However, the teacher could only move within the boundaries of the virtual classroom, and couldn't move out of the classroom. These 3DoF and 6DoF design could enhance the immersion of the remote class and improve the students' attention and interest in learning for a certain long time period.  **User story:**  Tom, Jerry, Emma, and Anna, four high school students, which are divided into two groups initially, are taking an English grammar course by Bob virtually. On the student side, all students can use their access equipments (VR/AR, smart phone, or any other possible device) to join the class and use headphone with head-tracking to hear the sound in the virtual classroom. If using smart phone to join the remote class, students could hold the smart phone to rotate their head, and the scene in the screen could change with the rotation (like a 360-degree video), also the correct sound could be heard in the corresponding video scene. The access equipment can receive and decode the bitstream of audio and video. Headphones can also be used for recording. On the teacher's side, Bob is in a conference room which has four displays with loudspeaker and a professional mic (or headphone only if under constrained conditions). The displays with loudspeaker can recreate the Audio-Visual or audio-only scene of students.  In this example, teleconferencing system constructs a virtual classroom with five participants, which are Bob, the English teacher, positioned at the front of the virtual space, and the four students are seated in front of Bob with two groups for group discussion.  In this Immersive and focused remote class, students can see and hear the teacher and other classmates. If the students don’t want to hear other classmates voice, they can turn off the sounds of others. The teacher can see and hear all the students to notice the students' activities in class.  The location of teacher and students in virtual space and their common activities are as follows:   1. 6DoF moving teacher: the teacher Bob could move in the virtual classroom, and his head could face different directions. Two ways could carry out the virtual moving of teacher. First, similar to video games, the teacher's moving can be carried out through a joystick. Second, obtaining the teacher's position in the physical classroom and then mirroring it to the virtual classroom. Compared with monotonous mono audio, immersive audio allows students to feel the changes in the teacher's voice naturally, which is helpful for reducing "zoom fatigue" and improving students' attention a bit. At the same time, it is a good way for the students to feel the teacher's speaking towards or backwards them, which is the same feeling as in a physical classroom.   In general, Bob would have four patterns in the space of classroom, and students could feel the difference among these four patterns:   * Close to blackboard and facing blackboard (Figure 1(a)). For example, Bob is writing new Grammar knowledge points and reading them simultaneously. Both students can hear the teacher is backing to them, but the listening experience of Group A and Group B is different. Because Bob is closer to Group A, the orientation angle between him and the two groups is also different. Group A will hear a louder voice in front of them, and Group B will hear the teacher's voice on their left. * Close to blackboard and facing students (Figure 1(b)). For example, Bob explains his writing on the blackboard to all students, Group A will hear his voice on their right, and Group B will hear his voice in front of them. Also, Bob can rotate his head to face different students based on the student's activities. During this process, students can hear whether the teacher is facing them or not. * Close to students and facing students (Figure 1(c)). For example, Tom and Jerry in Group A have a question to ask Bob, and the teacher moves close to them and faces them to solve their puzzles. In this situation, Group B will hear Bob is on the left side and not talk to them, and Group A will hear Bob on their right hand and face them to answer the question. * Close to students and facing blackboard (Figure 1(d)). For example, Emma in Group B is asked to answer the question written on the blackboard. The teacher would move close to Emma, watch the blackboard, and listen to her answer, then Bob comment on her answer in this position. Emma will hear Bob near her left hand and talking to her in this situation.     **Figure 1. Four space patterns of teacher status.**   1. 3DoF rotating students (suitable for VR/AR device). If there is no VR / AR device, students can also rotate to different directions through the handheld smartphone with the help of headtracking device: the students could not move in the classroom, but the orientation of their heads could change from time to time. In this way, the speakers can feel the discussion environment in the virtual classroom almost the same as in a physical classroom. Improving the immersion of remote class can prevent students from being tired or bored too easily. The orientation of students would have two patterns:  * Facing teacher (Figure 2(a)): all students stay unmoved and listen to the sound of teacher. Some usage stories have been described above. * Facing their teammate (Figure 2(b)): For example, during group discussion, Jerry is talking to Tom in Group A, and Emma is talking to Anna in Group B. Tom can hear Jerry's voice in front of him and realize that he is talking to himself. At the same time, Tom can also hear Emma's voice, but he can feel that Emma's voice is small and she is not talking to himself.     Figure 2. Two space patterns of student’s status.  Advantage: The above description of different user stories shows that teacher and students in class will have different location, and people can feel the differences from each state. Ordinary remote class is boring, making teachers and students get tired faster, however immersive classes can avoid this, while also help to improve students' concentration in learning and ensure that the teacher keeps track of the students' status. By simulating group discussions as in the physical classroom, students can hear the talks of other students in the virtual classroom, and this atmosphere helps to stimulate the enthusiasm of students to speak and discuss. |
| **Categorization** |
| **Type: <Mono, Stereo, Immersive>**  **Degrees of Freedom: <0DoF, 3DoF, 6DoF>**  **Delivery: <Conversational>**  **Media Components: <Audio-only, Audio-visual>**  **Device: <VR device, tablet, cellphone>** |
| **Preconditions** |
| Required:  The teacher Bob needs recording device, playback device and localization device.  Students need playback devices and device embedded with IMU sensors.  Potentially required:  VR/AR device for both teacher and students. |
| **QoS/QoE Considerations** |
| QoS: controlled network or prioritized fixed line should be used for high-quality conversation.  QoE: simple immersive audio rendering/binauralization quality for multiple parties, simulate the reverberation of the classroom. |
| **Feasibility** |
| Capture the spatial information: TWS headphones with embedded gyroscope sensors and spatial audio-capable are popular nowadays, thus it’s a valid assumption that measuring user’s orientation and location would be easier if necessary as the popularity of these type of devices grows.  Rendering: according to the user's listening equipment, the rendering mode can be flexible and diverse. Various spatial audio rendering technologies have been commercially available, such as HRTF, VBAP.  Data compression: according to the performance and bandwidth conditions of the user equipment, the audio data received by the user can be object-based or channel based. The relevant coding technology is relatively mature and can effectively compress the data of the channel or object.  The specific implementation is not limited to the content described above. Other solutions that can capture the spatial audio, encoding, decoding and rendering the audio data can also be used to implement this scenario. |
| **Potential Standardization Status and Needs** |
| Required:   * Support for high-quality encoding of immersive audio captured by smartphones (including smartphone accessories). * Spatial audio signal rendering.   Potentially required:   * Rendering scalability, i.e., a spatial signal is rendered at lower spatial dimension (e.g., as mono or stereo only). * a bit-stream could be channel-based, object-based or scene-based. |

1. **Conclusion**

The source provides a usage scenario and proposes incorporate it in the IVAS-9 P-doc.

# **References**

[1] Tdoc S4-200306: IVAS-9 Usage Scenarios v0.1.0\_clean.

[2] Tdoc SP-170611: New WID on EVS Codec Extension for Immersive Voice and Audio Services.

[3] Tdoc S4 220697: Audio solution for AR/VR usage scenario.

Annex A:  
Collection of IVAS Example Usage Scenarios

**Source: Editor[[1]](#footnote-1)**

**Title: IVAS [Example Usage Scenarios] (IVAS-X)**

**Version: 0.0.1**

1. **Scope**

This document collects example usage scenarios for the Immersive Voice and Audio Services (IVAS). The purpose of this collection is to create industry awareness of IVAS and to trigger interest at an early stage, even prior to IVAS standard finalization. After successful standardization and characterization, these examples should be incorporated into the IVAS TR.

1. **Introduction**

According to the IVAS codec WID [1] Immersive Voice and Audio Services are expected to cover UE-originated conversational and non-conversational use-cases, described in 3GPP TR 22.891 and TR 26.918. The study on eXtended Reality in 5G (XR5G) is expected to collect further compelling use casess that will require spatial audio and that may be enabled with the IVAS codec.

The following is a normalized description of IVAS example usage scenarios that has been collected using the usage scenario template provided in paragraph 3.0.

1. **IVAS Example Usage Scenarios**
   1. **Usage scenario template**

Table 3.0 Proposed Usage Scenario Collection Template

|  |
| --- |
| **Usage Scenario Name** |
| <add usage scenario name> |
| **Description** |
| <add detailed usage scenario description> |
| **Categorization** |
| **Type: <Mono, Stereo, Immersive, AR, VR, XR, MR>**  **Degrees of Freedom: <0DoF, 3DoF, 3DoF+, OD 6DoF, 6DoF>**  **Delivery: <Local, Streaming, Interactive, Conversational>**  **Media Components: <Audio-only, Audio-Visual>**  **Device: <UE, HMD, Glasses, Automotive, …>** |
| **Preconditions** |
| <provides conditions that are necessary to run the usage scenario, for example support for functionalities on the end device or network> |
| **Requirements and QoS/QoE Considerations** |
| <provides a summary on potential requirements as well as considerations on KPIs/QoE as well as QoS requirements> |
| **Feasibility** |
| <provides a summary on how the implementation of such a usage scenario using the IVAS codec is anticipated> |
| **Potential Standardization Status and Needs** |
| <identifies potential standardization needs> |

* 1. **Telephony Usage Scenarios**

The following is a collection of IVAS telephony usage scenarios.

* + 1. **Stereo and Immersive Telephony**

[TBD.]

* + 1. **VR Telephony**

[TBD.]

* 1. **Conferencing Usage Scenarios**

The following is a collection of IVAS conferencing usage scenarios.

* + 1. **Spatial conferencing**

[TBD.]

* + 1. **VR Conferencing**

[TBD.]

* + 1. **Virtual Meeting**

[TBD.]

* + 1. **Remote class participation**

[TBD.]

* + 1. **In-Game communications**

[TBD.]

* + 1. **XR Meeting**

[TBD.]

* + 1. **XR Convention / Poster Session**

[TBD.]

* 1. **User-generated content distribution Usage Scenarios**

The following is a collection of IVAS usage scenarios pertaining to user-generated content distribution.

* + 1. **Immersive and VR content distribution**

[TBD.]

# **References**

[1] Tdoc SP-170611: New WID on EVS Codec Extension for Immersive Voice and Audio Services.

1. Tbd. [↑](#footnote-ref-1)