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

**Title: Update Immersive and focused remote class participation usage scenario in IVAS Usage Scenarios (IVAS-9)**

**Document for: Discussion & Agreement**

**Agenda Item: 7.5**

**1. Introduction**

The source is an updated proposal for immersive and focused remote class participation usage scenario in IVAS Usage Scenarios (IVAS-9) [1]. The reversion is based on the discussion of SA4#120e meeting.

In summary, the edits address the following points: the meaning of 6DoF speaker and 3DoF speaker; what students could see and hear by using the head-tracking device; give more explanation about the moving of the 6DoF teacher; modify the references and some necessary corrections for the implementation.

**2. Detailed descriptions**

……

**3.2.4 Remote class participation**

**3.2.4.1 Immersive and focused remote class participation**

|  |
| --- |
| **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. 3DoF speaker(s), which can only change orientation in a virtual classroom, the position being fixed.
4. 6DoF speaker, able to change both position and orientation in the virtual classroom.

In this usage scenario, a teacher and one or more students may move freely in the physical environment, but the students can only interact as 3DoF speakers with predefined positions in the virtual classroom. These predefined positions in virtual classroom are similar to the seats in a physical classroom, which is initialled by the teacher. Students can choose the positions they want to sit in the class. But the positions cannot be changed once the class begins. A teacher may interact as a 6DoF speaker moving within the boundaries of the virtual classroom. The usage scenario aims to 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 equipment (VR (e.g. head mounted), AR (e.g. glasses), smartphone, 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 smartphone to join the remote class, students could hold the phone in front of them by their hands. When they rotate their heads, they can move the phone follow the rotation of head. The scene in the screen change follows the rotation (like a 360-degree video), also the correct sound is 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. They can choose classmates which they want to speak. So does the teacher.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, teacher can move physically if room available and the measured position can be used to mirror physical movement to the virtual classroom. In the second way, the size of virtual classroom should be the same as the teacher's physical classroom, the initialed position in the virtual space can be set by the teacher according to his initial physical position. Physical movement could enhance the experience of the teacher, but needs enough space and location equipment.

Compared with mono audio, immersive audio allows students to feel the changes of teacher’s location naturally, which is helpful for avoiding quick 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 similar feeling as in a physical classroom as much as possible. 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 back towards students (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 back towards students (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.

Logo, company name  Description automatically generated**Figure 1. Four space patterns of teacher status** (snapshots)**.**1. 3DoF rotating students (suitable for VR/AR device). the students cannot move in the virtual classroom, but the orientation of their heads in the virtual classroom could change from time to time. If there is no VR / AR device, students can also have an immersive experience with smartphone. For example, Jerry is facing the blackboard (Figure 2(a)), he could see the black board and the teacher and hear the teacher’s voice on his left front. Then he changes his orientation to the left and hold the smartphone in front of him, he could see student Tom and hear the teacher’s voice on his left back (Figure 2(b)). In this way, the students can feel the discussion environment in the virtual classroom almost the same as in a physical classroom, even there is no VR/AR device. 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 realizes that Jerry is talking to him. At the same time, Tom can also hear Emma's voice, but he can feel that Emma's voice is distant and she is not talking to him.

Graphical user interface  Description automatically generated with medium confidenceFigure 2. Two space patterns of student’s status (snapshots).Advantage: In ordinary remote classes, students heard the teacher's monotonous voice hours by hours and the concentration will gradually decrease, just like drivers would feel sleepy when they see the monotonous scenery all the time on the motorway. The 6DoF moving teacher and 3DoF rotating students could make the voices less monotonous and more immersive in the remote class. The above description of different user stories shows that teacher and students in class will have different location in the virtual classroom, and people can feel the differences from each state. Students could hear whether the teacher is facing them or turned away from them, being nearby or far away. In this way, teachers and students may get less tired than ordinary remote classes, and can improve students' concentration in learning. It may also help the teacher to better keep track of the students' learning. By simulating group discussions as in a physical classroom, students can hear the talks of other students in the virtual classroom, and this atmosphere may help 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/AR device, tablet, cellphone>** |
| **Preconditions** |
| Required:The teacher Bob needs recording device, playback device and localization device (tracking position and orientation).Students need recording device, playback devices, as well as device embedded with IMU sensors (tracking orientation) if any.Potentially required:VR/AR visual device for both teacher and students. Under this condition, the immersive experience of remote class is more natural. |
| **QoS/QoE Considerations** |
| QoS: controlled network or prioritized fixed line should be used for high-quality conversation.QoE: simple and practical immersive audio rendering/binauralization for multiple parties, simulate the reverberation of the classroom. |
| **Feasibility** |
| Capture the spatial information: TWS headphones with embedded gyroscope sensors, location technology and spatial audio-capable are popular nowadays, thus it’s a valid assumption that measuring user’s orientation and location would be easier.Rendering: depending on the user's listening equipment, the rendering can be adapted. Various spatial audio rendering technologies exist.Data compression: according to the performance and bandwidth conditions of the user equipment, the audio received by the user may for example be object-based or channel+objects representations. The channel (bed) could contain environmental information. Relevant coding technologies may be used to 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:* Efficient high-quality encoding of immersive audio, e.g. captured by smartphones (and/or smartphone accessories).
* Head-tracking information interface for decoder/renderer.
* Binaural rendering of immersive audio to headphones and spatial audio signal rendering to loudspeakers.

Potentially required:*
* TBD
 |

……

1. **Conclusion**

The contents are improved and supplemented in this update version. It is recommended to add these updates in IVAS-9 and remove the square brackets.

# **References**

[1] S4-221105 IVAS-9 Usage Scenarios v0.2.0\_cln.