**Title: Proposal on Microphone Array Signal Processing**

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

**Document for: Discussion & Agreement**

**Agenda Item: 7.8**

# Introduction

We have summarized the principle of acoustic design of immersive audio in chapter 6 of TR 26.933[1]. This proposal is to start the discussion of the microphone array signal processing in chapter 7 as clause 7.2. The subsequent updates will follow the current structure.

# Content

…

[

## Microphone Array Signal Processing

According to previous investigations, Microphone Array Signal Processing (MASP) is an essential step for stereo capture, parametric spatial audio capture and non-parametric spatial audio capture, the basic processing is to transform the raw microphone signals into an expected audio representation, and enhancement could be done if necessary.

### MASP on UE for Channel-based

#### MASP for Stereo

In Clause 4.1, current mobile phones with dimensions of approximately 15cm in length and 7cm in width. Furthermore, a minimum of two microphones is required for stereo audio capture. Consequently, it is logical to use the most basic microphone array configuration as an example: two microphones positioned at a distance of less than 15cm or 7cm, depending on the orientation of mobile phones.So this subchapter will focus on how to complete microphone array signal processing for producing stereo signals based on such microphone array.

### MASP on UE for Binaural

According to clause 6.2.1, binaural capture on UE appears to be the sole format capable of directly obtaining audio from the raw microphone signals via the earbuds' microphones, enhancement processing could be done for better performance.

### MASP for Scene-based

#### FOA

FOA signal model is very clear, it consists of four coincident signals: W, X, Y and Z, where W is an omnidirectional signal, while X, Y and Z are figure 8 directional signals aligned with the cartesian coordinate axes. So, the aim of the microphone array signal processing for FOA is to generate standard four coincident signals [2].

##### Matrix on current FOA microphones

The current FOA microphones utilize a tetrahedral configuration, which comprises four cardioid microphones oriented in the directions of Front-Left-Up (FLU), Front-Right-Down (FRD), Back-Left-Down (BLD), and Back-Right-Up (BRU). The W, X, Y, Z component are produced through matrix multiplying of the four cardioid signals, refer to equation(1). not follow the SN3D and ACN channel order.

#### HOA

HOA is for further study.

### MASP for MASA

MASA format signals consist of audio signals and metadata. The metadata refer to 26.250 are derived from analysis of microphone raw signals, so microphone array signal processing is an essential module for producing MASA signal.

### MASP for Object-based

According to 26.250 [3], Object-based audio consists of 1-4 individual mono object streams with associated metadata. Many existing technologies can be used to obtain object-based audio.

#### Mono object stream

The mono object stream may need audio with high quality and sufficient SNR, characteristics that closely match the existing mono audio solution. Consequently, the mono object stream may be derived from the current mono audio solution provided by UE.

#### Associated Object Metadata

A minimal set of object metadata associated is the object position in the polar coordinate system described using azimuth [-180°,180°] and elevation [-90°, 90°]angle.

Direction Of Arrival (DOA) is commonly utilized in current audio services, it is to determine the direction of the audio that needs to be processed. The direction information can also be set as associated object metadata to describe the position of one audio object.

]

…

# Conclusion

It is proposed to include section 2 into TR 26.933 as the chapter 7.2 Microphone Array Signal Processing.

**References**

1. S4-231944: [FS\_DaCED] TR 26.933 v0.3.0
2. Franz Zotter, Matthias Frank, “Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality”, Springer Nature, 2019
3. 3GPP TS 26.250: "Codec for Immersive Voice and Audio Services (IVAS); General overview