**Merged proposal:**

**Conclusions and Recommendations**

The outcome of this study is focused on immersive audio capture in mobile phones.

For playback use case, devices support stereo, and/or binaural audio playback through on-device loudspeakers and headphones. However, these devices still lack immersive capture functionality will widespread accelerate the, this may present a significant barrier to the widespread deployment of 3GPP IVAS. To significantly enhance the user experience and facilitate the faster deployment of 3GPP IVAS codec, the following points attempt to identify what may be needed for IVAS input from the raw microphone signals captured by devices from the sending point of view.

* Before supplying the raw multi-microphone captured signal, support for *equalization filter* design may be needed to achieve balanced frequency response characteristics across all microphones.
* As Category A device orientation can be changed during capturing, enhancement such as orientation adjustments may be needed.
* Support for converting multi-channel raw input to IVAS encoder input for headphone and/or loudspeaker consumption.
* When a binaural microphone is used for capturing, support may be needed for binaural to stereo conversion (for loudspeaker consumption at the receiver side)
* Support for multi-channel AEC, noise reduction and beamforming to improve the captured signal quality if needed.
* As the Category A devices are available in the different price range, at least support for stereo capturing may be recommended.
* Respective latencies for enhancement modules needs to be identified and defined.

So, aiming at opening the immersive audio market and accelerating the deployment of IVAS codec, an immersive audio capture solution is necessary, so, it is recommended to start a minimum performance implementation(s)~~normative~~ work targeting Category A devices which smartphones are the highest priority which take the above aspects into account.

Note:

1. Latency is a critical element but was not studied in the DaCED study phase.

2. Premium headphones often come with built-in processing capabilities, such as active noise cancellation (ANC), and other audio enhancement features. These features are processed directly within the headphones, which can reduce the need for additional processing in UE terminals. On the other hand, many other headphones, particularly those at lower price points, may not support or process these enhancements independently. As a result, the UE terminal may still need to perform some audio processing tasks to ensure that the audio quality meets certain requirements. So, a minimum quality may need to be defined for ensuring suitable captured signals.

3. UEs should be able to negotiate the best appropriate format depending on the capturing and rendering capabilities, and network environment characteristics. Support for signal negotiation needs to be defined.

 4. Methodology for checking the conformance of audio capture processing solutions from raw microphone signal to IVAS input shall be defined.

Original proposal from Xiaomi in S4-241443

## 9 Conclusion summary and recommendations

To effectively deploy IVAS codec, an end-to-end immersive audio solution is needed. Currently, stereo and binaural audio have already been deployed in receiving end by most UEs which have at least two speakers for stereo playback and can connect to earbuds for binaural audio playback.

It is different in the sending end, most UEs can hardly support immersive audio capture. The lack of immersive audio capture capability may presents a significant barrier to the widespread deployment of IVAS, so to provide the capability of immersive audio capture, we may significantly enhance the user experience and pave the way for the widespread deployment of IVAS codecs on market.

UE hardware and software hold the potential to deliver immersive voice and audio services with corresponding technologies. Most current UEs have more than 2 microphones, and can connect to earbuds for immersive audio playback. Among the UEs listed in clause 4, smartphones have become an indispensable part of people’s daily lives, not only as a means of real-time communication, but also as an important device for recording lives, creating contents and enjoying entertainments.

Aiming at opening the immersive audio market and accelerating the deployment of IVAS codec, an immersive audio capture solution is necessary, so, it is recommended to start a normative work of immersive audio capture for smartphones.

Original proposal from Panasonic in S4-241511

**10. Conclusions and Recommendations**

Different UE terminals and their corresponding designs discussed in the above clause are grouped into following categories:

* Category A: Mobile, Tablets, Laptops, AR Glasses, Headphones
* Category B: Watches
* Category C: Car.

The market need for immersive voice communication (more than mono) for Category B terminal is unknown. if Category B is used as a peripheral device, it can be considered a part of Category A. Additionally, the internal acoustics of Category C are entirely different from those of Categories A and B. Sufficient study on Category C was not captured in this TR, so it can be deferred for future study.

Modern Category A devices (Mobile, Tablets, Laptops, AR Glasses, Headphones) have multi-microphone setups, but they still do not support immersive voice and audio capturing for communication. Given that the 3GPP IVAS codec is available for market adoption, the need for supporting Immersive voice and audio capturing for IVAS must be identified. The following points attempt to identify what may be needed for IVAS input from the raw microphone signals captured by Category A devices.

IVAS supports multiple audio formats, such as stereo, multi-channel, Independent Streams with Metadata (ISM), Scene-based Audio (SBA), Metadata Assisted Spatial Audio (MASA), and Binaural. Given that Category A devices are mostly used in real-world environments, raw multi-channel microphone capture signals may not be suitable as input to the IVAS codec. As such, following are identified to missing in the Category A devices.

* Before supplying the raw multi-microphone captured signal, support for equalization filter design may be needed to achieve balanced frequency response characteristics across all microphones.
* As Category A device orientation can be changed during capturing, enhancement such as orientation adjustments may be needed.
* Support for converting multi-channel raw input to binaural and/or stereo, especially for headphone and/or stereo loudspeaker consumption.
* When a binaural microphone is used for capturing, support may be needed for binaural to stereo conversion (for loudspeaker consumption at the receiver side)
* Support for multi-channel AEC, noise reduction and beamforming to improve the captured signal quality.
* Conversion of raw multi-channel input to MASA supported format for IVAS MASA voice communication
* As the Category A devices are available in the different price range, at least support for stereo capturing may be recommended.
* End–to–End Latency and respective latencies for enhancement modules needs to be identified and defined.

Note:

1. Latency is a critical element but was not studied in the DaCED study phase.

2. Premium headphones often come with built-in processing capabilities, such as active noise cancellation (ANC), and other audio enhancement features. These features are processed directly within the headphones, which can reduce the need for additional processing in UE terminals. On the other hand, many other headphones, particularly those at lower price points, may not support or process these enhancements independently. As a result, the UE terminal may still need to perform some audio processing tasks to ensure that the audio quality meets certain requirements. So, a minimum quality may need to be defined for ensuring suitable captured signals.

3. UEs should be able to negotiate the best appropriate format depending on the capturing and rendering capabilities, and network environment characteristics. Support for signal negotiation needs to be defined.