|  |
| --- |
| 3GPP TR 26.933 V0.0.3 (2023-04) |
| Technical Report |
| 3rd Generation Partnership Project;Technical Specification Group SA; Study on Diverse Audio Capturing system (Release 19)  |
|   |
|  |  |
|  |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented.This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification.Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. |

|  |
| --- |
|  |
| ***3GPP***Postal address3GPP support office address650 Route des Lucioles - Sophia AntipolisValbonne - FRANCETel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16Internethttp://www.3gpp.org |
| ***Copyright Notification***No part may be reproduced except as authorized by written permission.The copyright and the foregoing restriction extend to reproduction in all media.© 2022, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).All rights reserved.UMTS™ is a Trade Mark of ETSI registered for the benefit of its members3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational PartnersLTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational PartnersGSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 4

Introduction 5

1 Scope 6

2 References 6

3 Definitions of terms, symbols and abbreviations 7

3.1 Terms 7

3.2 Symbols 7

3.3 Abbreviations 7

4 Factors of different UE categories related to audio capture 7

5 Components used in audio capture 7

6 Acoustic design 8

7 Signal processing 8

8 Example audio capture processing solutions 8

9 Conclusions and Recommendations 8

Annex <X> (informative): Change history 9

# Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# Introduction

Providing immersive voice and audio services by end-user devices is becoming more and more practicable with the development of 4G/5G technologies. Related requirements have been investigated in 3GPP TR 22.891. Several use cases for VR are envisioned in TR 26.918, and for these cases the corresponding audio capturing system are generally considered. As such, capturing capability is crucial for making truly immersive voice and audio experiences.

Due to physical constraints on their outline shapes and sizes, the end-user devices are usually configured with different numbers of microphones and also different microphone setup configurations, hence different audio capturing capabilities are expected. Based on this, the present document gives diverse audio capturing system.

# 1 Scope

This document addresses audio capturing configurations for end-user devices, which is to make the devices to have audio capturing capability in order to provide truly immersive voice and audio service.

[

The agreed objectives are as following:

* Study audio capturing configurations for end-user devices considering:
	+ Different UE form factor designs and categories such as:
		- Smartphone, Headset, Earbud, AR/VR device, Vehicle, Desktop, Laptop, Conference phones and other envisioned future devices.
	+ Microphone placement/orientation/quantity and device positioning, acoustic structures coupled to the microphones, and necessary processing for audio format.
	+ Analog-to-Digital conversion if needed, Signal to Noise Ratio aspects, frequency band & group delay compensation.
	+ Methods for improving the immersive audio experience, including signal enhancement (e.g. gain, noise, echo and spatiality control).
	+ Test methods for characterizing the audio capture performance from e.g., :
		- ITU-T P.800/811, ITU-R.BS2132, ITU-R BS.2051-3, ATIAS
	+ Different device tiers (low-end, middle-end, high-end)
	+ Modelling and theoretical analysis for capture configuration.
	+ The set of supported audio formats generated by the different audio capturing configurations, including mono, stereo, binaural, MASA, multichannel, object-based audio and scene-based audio.
* Study example audio capture processing solutions that can be used in conjunction with immersive voice and audio services codecs.

]

Editor’s Note: the scope is for further detailed based on the objectives and input contribution.

# 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 26.891: "5G enhanced mobile broadband; Media distribution".

[3] 3GPP TR 26.918: "Virtual Reality (VR) media services over 3GPP".

[4] 3GPP TS 26.119: "Media Capabilities for Augmented Reality".

[...] ……

# 3 Definitions of terms, symbols and abbreviations

This clause and its three subclauses are mandatory. The contents shall be shown as "void" if the TS/TR does not define any terms, symbols, or abbreviations.

## 3.1 Terms

For the purposes of the present document, the terms given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

Definition format (Normal)

**<defined term>:** <definition>.

**example:** text used to clarify abstract rules by applying them literally.

## 3.2 Symbols

For the purposes of the present document, the following symbols apply:

Symbol format (EW)

<symbol> <Explanation>

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

VR Virtual Reality

[

# 4 Factors of different UE categories related to audio capture

Editor’s Note:

* *Collect relevant information on potential UEs like smartphone, headphone, XR glasses, etc.*
* *The shape, structure of UE.*
* *Available computer power according to current device and tendency.*

## 4.1 Structure Size

Since 2012, the structure size of mobile phones has been rising in length, width, and thickness, indicating a continuous increase in mobile phone size. This may be in response to the strong demand from consumers for multimedia and gaming functions on their phones, as well as the increasing requirements for microphone and camera quantity and battery consumption. The evolution trend of mobile phones is towards full-screen, which provides a market foundation for the increase in screen size.

The detailed data of structure size are listed in Appendix A

### 4.1.1 Length

The length of mobile phones has gradually increased from 12.38cm in 2012 to 16.88cm in 2022, with an average length of 15.26cm, according to the investigation, showing an upward trend. With the development of mobile phone models, some phones are no longer limited to the 16:9 aspect ratio, e.g., there are now styles with 18.5:9 and 19.5:9. Although high aspect ratio screens can display more information, most video contents are still in the traditional 16:9 format, so too high an aspect ratio is not conducive to the video display.

The tendency of mobile phone length in recent years is shown in Figure 4.1.1-1



**Figure 4.1.1.1 The tendency of mobile phone length**

### 4.1.2 Width

According to the mobile phone data surveyed, the width of mobile phones was around 5.86cm in 2012, while in 2022, the width of mobile phones was changed to around 7.55cm, with a maximum value of 8.06cm. In recent years, the average width of mobile phones has been 7.32cm, and an increment in the length of mobile phones generally follows an increment in width. This is a reasonable evolution tendency for the purpose of function requirements and appearance

The tendency of mobile phone width in recent years is shown in Figure 4.1.2-1



**Figure 4.1.2-1 The tendency of mobile phone width**

### 4.1.3 Depth

Among the phones investigated, the thinnest one measures 0.64cm, the thickest one measures 0.992cm, and the average thickness is 0.81cm.

The tendency of mobile phone width in recent years is shown in Figure 4.1.3-1



**Figure 4.1.3-1 The tendency of mobile phone depth**

### 4.1.4 Summary

According to the investigations, the summary is as follows:

* The maximum values of length, width, and height are 168.78mm,80.6mm and 9.92mm.
* The minimum values are 123.8mm,58.6mm, and 6.4mm.
* The average values are 152.65mm,73.17mm and 8.08mm.
* The 95% Confidence Interval (CI) are (149.60 mm,155.69 mm), (71.92 mm,74.42 mm) and (7.85 mm,8.31 mm).

Editor’s Note: this is basis for further work

# 5 Components used in audio capture

Editor’s Note:

* Documentation of components may be used in diverse audio capture.
* Relevant components like microphone, AD converter, etc.

## 5.1 Component

### 5.1.1 Microphone

The function of microphone is to convert sound pressure signal to analog electrical signal in circuit.

4 types of microphones popular in the market are described in this proposal. These microphones have unique advantages in UE's immersive audio. They are classified to dynamic microphone, condenser microphone, Micro-Electro-Mechanical Systems (MEMS), contact microphone.

#### 5.1.1.1 Dynamic microphone

Dynamic microphone is one of the popular microphones on market. The most advantage of dynamic microphone for UE is it doesn’t need for external power; the entire recording system will be easier. Another advantage is durability, make it more suitable for loud and high-pressure situation. But it usually has a disadvantage that it is less sensitive to high frequencies.

Dynamic microphone uses a small movable induction coil, which positioned in the magnetic field and is attached to the diaphragm. The current signal generates when the movement of the diaphragm causes the coil to also move within a magnetic field.

#### 5.1.1.2 Condenser microphone

Condenser microphone is another popular microphone on market, especially for immersive audio. Most immersive system is using condenser microphones, like ambisonic microphone and external stereo microphone for mobile phone. It’s popular for its high sensitivity, wide frequency response, low noise. However, the condenser microphone requires a power source, and in the case of most professional condenser microphones, it specifically requires 48V phantom power. Meeting this requirement can be challenging for UE device consider the channel number of immersive audio.

Condenser microphone uses capacitor to convert sound waves to electrical signal. The capacitor consists of two plates, one of them is a diaphragm that vibrates in response to sound waves. The diaphragm vibrates and changes the distance between the two plates. Then the capacitance changes which influences the electrical signal.

#### 5.1.1.3 Micro-Electro-Mechanical Systems microphone

In the past decades, microphone for UE has change from carbon microphones to electret condenser microphones. Recently the MEMS microphone is spread rapidly, benefited from its advantages of high stability and small volumes.

According to the techniques of microfabrication, the MEMS microphone is much smaller and allow integrate other components including preamps, ADC with transducer in one package under the control of integrated microelectronics.

Which means for manufacturers, it much easier to build the capture system, MEMS microphone can output the digital signal directly. In other hand, it allows need to select the component more carefully. Since the microphone is much smaller and very uniform in their mechanical properties, it's suitable for UE and make immersive audio become possible for economic portable UE like mobile phone.

#### 5.1.1.4 Contact microphone

Contact microphone is a type of microphone that senses solid vibrations through direct contact with a surface.

Compared to the acoustic microphones, the contact microphones have the benefit of not to capture sound waves in the air, but to capture mechanical vibrations of the target object. Hence, it’s resistant to noise in air.

Nowadays, bone conduction microphone, which is a special kind of contact microphone, is very popular on TWS headphones. It is used to capture high SNR speech signal even in complex scenarios.

#### 5.1.1.5 Other microphones

TBD

5.1.1.6 Summary

From a size perspective, the MEMS microphones are the best choice for most portable UE (like mobile phone, headphone). The study will mainly focus on this miniature microphone consider the immersive audio system is much more complex.

Other microphones will also be considered, like the dynamic microphone and condenser microphone still dominate the professional audio industry.

## 5.2 Preamps

TBD

## 5.3 ADC

TBD

## 5.4 Clock

TBD

## 5.5 Directivity

Directivity is a very important part in immersive audio, every immersive audio format has requirement on directivity. Even for objective audio, we also need take care of the directivity to avoid the influence of environment noise.

### 5.5.1 Traditional approaches used in immersive audio

#### 5.5.1.1 Directional microphone capsule

Most directional microphone is using two closely diaphragms that electrically subtracted from each other to provide a range of polar patterns.

#### 5.5.1.2 Interference tube

Interference tube is usually used on shotgun microphones. Make it the more directional than a typical cardioid or supercardioid microphone.

Interference tube is a long, narrow extended tube that is placed in front of the microphone capsule and has multiple small holes along its length. It creates phase shfit for sounds arriving from off-axis directions, the off-axis sound will arrive at the diaphragm with varying phase relationships and so partially cancel one another out.



**Figure 5.5.1.2-1 The schematic diagram of interference tube**

#### 5.5.1.3 Binaural acoustic stimulation

TBD

### 5.5.2 Beamforming microphone array

Research on microphone array beamforming began in the late 1960s, although some basic principles can be traced back to the 1930s when directional microphones were invented. Early work in this field was strongly influenced by sensor array theory developed in the radar and sonar fields.

Beamforming is a very popular technology to achieve target directivity, though it’s mostly used for mono speech now, it is great potential in immersive audio. There are also many studies in this area.

This proposal starts with two fundamental technologies: Delay-sum and differential. And aim for the suitable solution for immersive audio on UE.

#### 5.5.2.1 Delay-sum microphone array



**Figure 5.5.2.1-1 The diagram of Delay-sum microphone array**

The basic idea of this technology is to delay the output of each microphone by an appropriate amount of time. The phase relationship between the microphones is carefully controlled to ensure that the signal form desired direction have the same phase so that they can be reinforced.

Though the delay-sum microphone array can obtain a very sharp directionality. However, the biggest problem with this beamformer is that its beam pattern changes significantly with frequency.

#### 5.5.2.2 Differential microphone array



**Figure 5.5.2.2-1 The diagram of Differential microphone array**

In Differential Microphone Array (DMA), the signals from two or more microphones are subtracted from each other to create a special directivity. The traditional directional microphone can also be seen as a special kind of differential beamforming.

By adjusting the weight and phase of the differential signal, we can all get different directivity like: cardioid, bidirectional (Figure-8), supercardioid, hypercardioid, subcardioid (wide cardioid).

Due to the smaller spacing between microphones, the size of array is usually smaller, making it easy to integrate into UE such as earphones, mobile phones, etc. Another characteristic of DMA is that its directivity is frequency-invariant; therefore, they are suitable for processing broadband speech and audio signals.

Editor’s Note: this is basis for further work

# 6 Acoustic design

Editor’s Note

* Relevant acoustic design content is envisioned.
* Including acoustic structure, microphone array design, etc.

## 6.1 Stereo microphone configurations

Stereo microphones can generally be classified as spaced, near-coincident, or coincident based on the angle and distance between the microphones. However, due to the current size of mobile phones, we will only list microphone configurations that distance between microphone is less than18cm, which is possible to install on mobile phone.

### 6.1.1 Near-Coincident

Near-Coincident using two directional microphones placed close together with an angle is to capture stereo audio. This configuration utilizes the angle and distance between the microphones to create a suitable time and level inter-channel difference.

#### 6.1.1.1 ORTF

ORTF stereo microphone uses two cardioid microphones with 17cm distance and 110° angle.



**Figure 6.1.1.1-1 The configuration of ORTF stereo microphone**

### 6.1.2 Baffled

A baffled configuration is a stereo recording technique that utilizes an acoustic baffle to increase the separation between the left and right audio channels. The baffle is typically a physical barrier that is placed between the two microphones.

#### 6.1.2.1 OSS (Optimal Stereo System)

This method utilizes a specially designed 30-cm disk covered with foam, with two omni-directional microphones mounted on opposite sides of the disk and angled slightly outward at 20°. The capsules of the two microphones are positioned 16.5 cm apart.



**Figure 6.1.2.1-1 The configuration of OSS stereo microphone**

### 6.1.3 Coincident

A coincident stereo microphone consists of two directional microphones placed at an appropriate angle at the smallest-possible spacings. Therefore, sound arrives with equal delay and different level and phase at microphones.

#### 6.1.3.1 X/Y

X/Y stereo microphone is commonly using two cardioid microphones ranging from 90-135°.



**Figure 6.1.3.1-1 The configuration of X/Y stereo microphone**

#### 6.1.3.2 Blumlein

Blumlein stereo microphone consists of two bidirectional (figure-eight) microphones with 90° angle at the same place.

****

**Figure 6.1.3.2-1 The configuration of Blumlein stereo microphone**

#### 6.1.3.3 M/S

M/S (mid-side) stereo microphone using a microphone (usually a cardioid) pointed forward, and a bidirectional (figure-eight) microphone oriented perpendicular to the directional microphone. The figure-eight microphone captures mid signal, and the cardioid microphone capture side signal. Therefore, we can obtain the left and right channel signal through the simple addition and subtraction.

In addition, controlling the ratio of the two signals, different angles can be obtained.



**Figure 6.1.3.3-1 The configuration of M/S stereo microphone**

### 6.1.4 Spaced

The spaced stereo microphone, also known as A/B stereo, is a stereo microphone technique that involves placing two omnidirectional microphones some distance apart from each other. This technique is commonly used with microphone spacings ranging from 0.6-1 meter.

The spaced stereo microphone technique utilizes the distance between two microphones to create a time difference and level difference between the left and right channels. This is caused by the difference in arrival time of sound waves at each microphone, as well as the absorption of sound by the air between the microphones.

NOTE: As most classic spaced configurations involve microphone distances greater than 30cm, which exceeds the size of current mobile phones, this aspect can only be listed for further study.

# 7 Signal processing

Editor’s Note

* Relevant signal processing content is envisioned
* Including relevant processing for audio format, enhancement solution for immersive, speech enhancement, etc.
* Relevant characterization of the audio capture performance.

# 8 Example audio capture processing solutions

Editor’s Note

* Example solutions can be guidance on usage in conjunction with immersive voice and audio services codecs.

# 9 Conclusions and Recommendations

Editor’s Note

* Provide recommendation on potential work for audio capturing based on the findings in this study.

]

Editor’s Note: the chapter structures are for further update.

Annex <A> (informative):

*Start each annex on a new page.*

*Annexes are labelled A, B, C, etc. and designated "informative".*

[

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Date (MMM-YY) | Length(mm) | Width(mm) | Depth(mm) |
| 1 | Mar-23 | 146.7 | 71.5 | 7.8 |
| 2 | Mar-23 | 160.8 | 78.1 | 7.8 |
| 3 | Mar-23 | 162.9 | 76.7 | 8.77 |
| 4 | Mar-23 | 162.9 | 76.7 | 8.77 |
| 5 | Dec-22 | 164.07 | 74.53 | 9.34 |
| 6 | Dec-22 | 164.35 | 75.29 | 9.7 |
| 7 | Sep-22 | 162.1 | 75.5 | 8.5 |
| 8 | Sep-22 | 162.1 | 75.5 | 9.92 |
| 9 | Apr-22 | 168.78 | 80.31 | 8.37 |
| 10 | Apr-22 | 163.7 | 73.9 | 8.5 |
| 11 | Mar-22 | 163.7 | 73.9 | 8.5 |
| 12 | Mar-22 | 160.3 | 72.6 | 8.7 |
| 13 | Sep-21 | 146.7 | 71.5 | 7.65 |
| 14 | Sep-21 | 163.6 | 74 | 9.1 |
| 15 | May-21 | 163.6 | 74 | 8.26 |
| 16 | Mar-21 | 163.6 | 74 | 8.7 |
| 17 | Mar-21 | 163.6 | 74 | 8.26 |
| 18 | Oct-20 | 146.7 | 71.5 | 7.4 |
| 19 | Oct-20 | 162.9 | 75.5 | 9.1 |
| 20 | Apr-20 | 158.2 | 72.6 | 8.95 |
| 21 | Dec-19 | 160.3 | 74.3 | 7.96 |
| 22 | Sep-19 | 150.9 | 75.7 | 8.3 |
| 23 | Apr-19 | 156.6 | 74.3 | 9 |
| 24 | Sep-18 | 143.6 | 70.9 | 7.7 |
| 25 | Sep-18 | 157.68 | 74.06 | 7.47 |
| 26 | Jun-18 | 157.91 | 74.27 | 7.48 |
| 27 | Apr-18 | 149.1 | 70.8 | 7.65 |
| 28 | Mar-18 | 155.1 | 75.2 | 7.4 |
| 29 | Oct-17 | 143.6 | 70.9 | 7.7 |
| 30 | Oct-17 | 150.5 | 77.8 | 8.2 |
| 31 | Sep-17 | 138.4 | 67.3 | 7.3 |
| 32 | Sep-17 | 165.32 | 80.09 | 7.45 |
| 33 | Jun-17 | 142.2 | 68.9 | 8.9 |
| 34 | Nov-16 | 150.5 | 77.8 | 8.2 |
| 35 | Oct-16 | 153 | 74.3 | 6.58 |
| 36 | Sep-16 | 138.3 | 67.1 | 7.1 |
| 37 | May-16 | 146.8 | 72.6 | 7.5 |
| 38 | Nov-15 | 157.1 | 80.6 | 7.9 |
| 39 | Oct-15 | 151.3 | 76.3 | 8.15 |
| 40 | Aug-15 | 141.6 | 71.2 | 7.8 |
| 41 | Jun-15 | 143.2 | 71.9 | 8.5 |
| 42 | May-15 | 147.9 | 73.45 | 6.44 |
| 43 | Apr-15 | 144.9 | 71.9 | 6.4 |
| 44 | Oct-14 | 158.1 | 77.8 | 7.1 |
| 45 | Sep-14 | 138.1 | 67 | 6.9 |
| 46 | Jul-14 | 139.6 | 69.7 | 7.5 |
| 47 | Jul-14 | 139.2 | 68.5 | 8.9 |
| 48 | Mar-14 | 152.6 | 75 | 9.2 |
| 49 | Sep-13 | 123.8 | 58.6 | 7.6 |
| 50 | Sep-12 | 123.8 | 58.6 | 7.6 |

# ]

Editor’s note: this is basis for further work.

Annex <X> (informative):
Change history

|  |
| --- |
| **Change history** |
| **Date** | **Meeting** | **TDoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
| 2023-02 | SA4#122 | S4-230317 |  |  |  | Initial version | V0.0.1 |
| 2023-04 | SA4#123-e | S4-230551 |  |  |  | Updated version based on SA4-post 122 24,March ,2023 | V0.0.2 |
| 2023-04 | SA4#123-e | S4-230646 |  |  |  | Update style and include agreed content in S4-230522 and S4-230523 | V0.0.3 |