|  |  |
| --- | --- |
| 3GPP TR 26.933 V0.0.1 (2023-02) | |
| 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 address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  http://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 members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® 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.*

# 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.

# 6 Acoustic design

Editor’s Note

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

# 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".*

Annex <X> (informative):  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change history** | | | | | | | |
| **Date** | **Meeting** | **TDoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
| 2023-02 | SA4#122 |  |  |  |  |  | V0.0.1 |