ETSI TS 126 260 V15.2.0 (2020-03)

TECHNICAL SPECIFICATION

5G;
Objective test methodologies for the evaluation of immersive audio systems
(3GPP TS 26.260 version 15.2.0 Release 15)

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/59af95de-97a1-4c33-b74c-0fcb1c2e78fa/etsi-ts-126-260-v15.2.0-2020-03

A GLOBAL INITIATIVE

Reference
RTS/TSGS-0426260vf20

Keywords
5G

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Copyright Notification

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.
The content of the PDF version shall not be modified without the written authorization of ETSI.
The copyright and the foregoing restriction extend to reproduction in all media.

© ETSI 2020.
All rights reserved.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members.
3GPP™ and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.
oneM2M™ logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners.
GSM® and the GSM logo are trademarks registered and owned by the GSM Association.

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Legal Notice

This Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/59af95de-97a1-4233-b74c-0fcb1c3e7efa/etsi-ts-126-260-v15.2.0-2020-03

Foreword

This Technical Specification 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.

Introduction

Audio is a key component of an immersive multimedia experience and 3GPP systems are expected to deliver immersive audio with a high Quality of Experience. However, industry agreed methods to assess the Quality of Experience for immersive audio are relatively few and the present document seeks to address this gap by providing objective test methods for the assessment of immersive audio.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/59af95de-97a1-4c33-b74c-0fcb1c2e78fa/etsi-ts-126-260-v15.2.0-2020-03

1 Scope

The present document specifies objective test methodologies for 3GPP immersive audio systems including channel based, object based, scene-based and hybrids of these formats. The subjective evaluation methods described in the present document are applicable to audio capture, coding, transmission and rendering as indicated in their corresponding clauses.

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] J. Fliege und U. Maier: "A two-stage approach for computing cubature formulae for the sphere," Dortmund University, 1999.

[3] ISO 3745 - Annex A: "Acoustics - Determination of sound power levels and sound energy levels of noise sources using sound pressure -- Precision methods for anechoic rooms and hemi-anechoic rooms - Annex A: General procedures for qualification of anechoic and hemi-anechoic rooms".

[4] ISO 1996 Acoustics: "Description, measurement and assessment of environmental noise".

[5] ANSI S1.4: "Specifications for Sound Level Meters".

[6] ISO 3: "Preferred numbers – Series of preferred numbers".

[7] B. Rafaely, “Analysis and design of spherical microphone arrays,” IEEE Transactions on Speech and Audio Processing, no. 13, 2005, pp. 135 – 143

iTeh STANDARD PREVIEW
(standards.iteh.ai)

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions 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].

spherical coordinates: The coordinate system used in this document is defined such that the x-axis points to the front, the y-axis to the left and the z-axis to the top (see Figure 0). Spherical coordinates are the distance r from the origin, the azimuth ϕ in mathematical positive orientation (counter-clockwise) and the elevation angle θ relative to the z-axis (with 0 degrees pointing to the equator and +90 degrees pointing to the North pole).

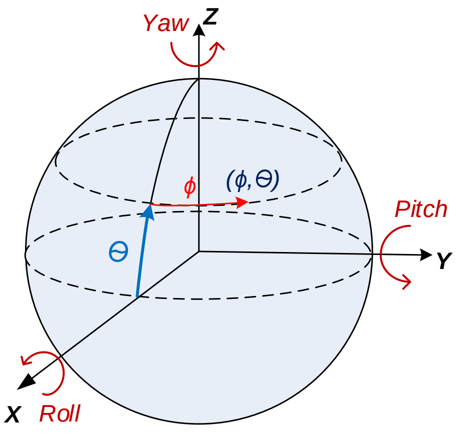


Figure 0: Spherical coordinate system

dBFS: dB full-scale, where 0 dBFS refers to the RMS level of a DC-free sinusoidal signal exercising the full scale of the digital interface/file.

3.2 Symbols

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

LA_{eq}	the sound level in decibels equivalent to the total A-weighted sound energy measured over a stated period of time.
ϕ	azimuth
θ	elevation

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].

iTeh STANDARD PREVIEW (standards.iteh.ai) Full standard: https://standards.iteh.ai/catalog/standards/sist/59af95de-97a1-4c33-b4ec-fcb1c2e78fa/etsi-ts-126-260-v15.2.0-2020-03

4 Objective Test Methodologies for Immersive Audio Systems

4.1 Objective Test Methodologies for Assessment of Immersive Audio Systems in the Sending Direction

4.1.1 Diffuse-field Send Frequency Response for Scene-based Audio

4.1.1.1 Introduction

This test is applicable to UEs capturing scene-based audio (e.g. First and Higher Order Ambisonics).

NOTE: Currently, the test method uses a periphonic loudspeaker array for generation of a diffuse-field. Additional loudspeaker setups for the derivation of the diffuse sound field are under consideration.

General test conditions

Free-field propagation conditions

- The test environment shall contain a free-field volume, wherein free-field sound propagation conditions shall be observed.
- The free-field sound propagation conditions shall be observed down to a frequency of 200 Hz or less.
- Qualification of the free-field volume shall be performed using the method and limits for deviation from ideal free-field conditions described in [3].

Test environment noise floor

Within the *free-field volume*, the equivalent continuous sound level of the test environment in each 1/3rd octave band, $L_{eq}(f)$, shall be less than the limits of the NR10 curve, following the noise rating determination procedures in [4].

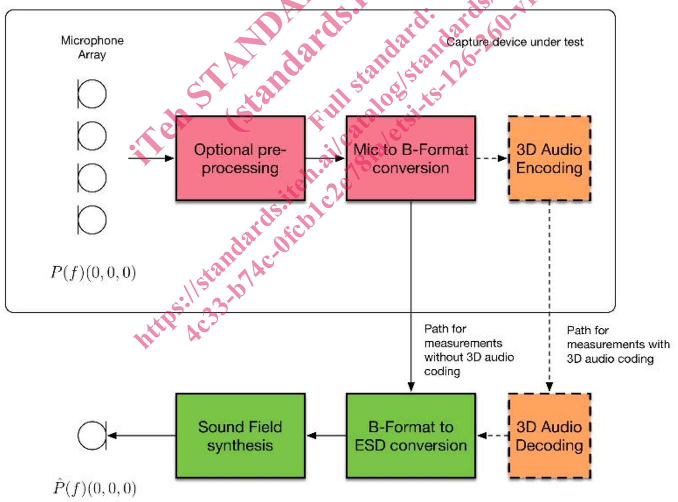
4.1.1.2 Definition

The Diffuse-field Send Frequency Response for Scene-based Audio is defined as the transfer function, $G(f)$, between:

$\hat{P}(f)$, the estimated sound pressure magnitude spectrum obtained from a diffuse-field scene-based audio capture and reference synthesis at the geometric center of a *free-field volume*; and

$a)P(f)$, the sound pressure magnitude spectrum obtained from a diffuse-field microphone recording the same diffuse field at the origin of a spherical coordinate system.

Figure 1 describes a typical block diagram for the scene-based audio sending direction with measurement points when using a periphonic loudspeaker array.



iTeh STANDARD PREVIEW (standards.iteh.ai) Full standard: https://standards.iteh.ai/catalog/standards/sist/59af95de-97a1-4c33-b74c-0fcb1c2e78f8/etsi-ts-126-260-v15.2.0-2020-03

Figure 1: Scene-based audio capture block diagram for sending direction measurements

Definition of Equivalent Spatial Domain

The equivalent spatial domain representation, **w**(t), of a N^{th} order Ambisonics soundfield representation **c**(t) is obtained by rendering **c**(t) to K virtual loudspeaker signals $w_j(t)$, $1 \le j \le K$, with $K = (N+1)^2$. The respective virtual loudspeaker positions are expressed by means of a spherical coordinate system, where each position lies on the unit sphere, i.e., a radius of 1. Hence, the positions can be equivalently expressed by order-dependent directions $\boldsymbol{\Omega}_j^{(N)}=(\theta_j^{(N)}, \phi_j^{(N)})$, $1 \le j \le K$, where $\theta_j^{(N)}$ and $\phi_j^{(N)}$ denote the inclinations and azimuths, respectively. These directions are defined according to [2] and reproduced in Annex B for convenience.

The rendering of $\boldsymbol{c}(t)$ into the equivalent spatial domain can be formulated as a matrix multiplication:

$$\mathbf{w}(t) = (\boldsymbol{\Psi}^{(N,N)})^{-1} \cdot \mathbf{c}(t),$$

where $(\cdot)^{-1}(\cdot)^{-1}$ denotes the inversion.

The matrix $\mathbf{\Psi}^{(N,N)}$ of order N with respect to the order-dependent directions $\boldsymbol{\Omega}_j^{(N)}$ is defined by:

$$\mathbf{\Psi}^{(N,N)} := [\mathbf{S}_1^{(N)} \quad \mathbf{S}_2^{(N)} \quad \ldots \quad \mathbf{S}_K^{(N)}],$$

with:

$$\mathbf{S}_j^{(N)} := [S_0^0(\boldsymbol{\Omega}_j^{(N)}) \quad S_{-1}^{-1}(\boldsymbol{\Omega}_j^{(N)}) \quad S_{-1}^{0}(\boldsymbol{\Omega}_j^{(N)}) \quad S_{-1}^{1}(\boldsymbol{\Omega}_j^{(N)}) \quad S_{-1}^{1}(\boldsymbol{\Omega}_j^{(N)}) \quad \ldots \quad S_N^N(\boldsymbol{\Omega}_j^{(N)})]^T,$$

where $S_n^m(\cdot)$ represents the real valued spherical harmonics of the order n and degree m.

The matrix $\mathbf{\Psi}^{(N,N)}$ is invertible so that the HOA representation $\boldsymbol{c}(t)\boldsymbol{c}(t)$ can be converted back from the equivalent spatial domain by:

$$\mathbf{c}(t) = \mathbf{\Psi}^{(N,N)} \cdot \mathbf{w}(t)$$

4.1.1.3 Test method with periphonic array

4.1.1.3.1 Test Conditions

Periphonic loudspeaker array

a) A *periphonic loudspeaker array* shall be placed within the free-field volume with the geometric center of the *periphonic loudspeaker array* coinciding with the geometric center of the free-field volume.

b) The *periphonic loudspeaker array* shall have a radius greater or equal than 1 meter.

c) The *periphonic loudspeaker array* shall be composed of $(N+1)^2$ coaxial loudspeaker elements. Each of the $(N+1)^2$ coaxial loudspeaker elements shall be equalized (if necessary) and level compensated to conform with the operational room response curve limits given in [5] Section 8.3.4.1. N should be equal or greater than the maximum ambisonics order supported by the device under test (DUT), e.g. N>=4 for a DUT supporting 4th order Ambisonics capture.

d) The $(N+1)^2$ coaxial loudspeaker elements shall be positioned according to the azimuth and elevation coordinates given in Annex B.

e) All coaxial loudspeaker elements shall be oriented such that their acoustic axis intersects at the geometric center of the *free field volume*.

f) The radius of each coaxial loudspeaker element shall be such that, at the geometric center of the *free-field volume*, the far field approximation for the coaxial loudspeaker axial pressure amplitude decay holds true.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/59a95de-57af-4e33-b74c-0fcb1c2e78fa/etsi-ts-126-260-v15.2.0-2020-03

4.1.1.3.2 Measurement

Reference Spectrum measurement for periphonic loudspeaker array method

a) A diffuse-field / random incidence, or multi-field microphone is mounted in the *free-field volume* such that the tip of the microphone corresponds to the geometric center of the *free-field volume* and the geometric center of the *periphonic loudspeaker array*.

NOTE 1: Diffuse-field / random incidence microphones, are described in [5].

b) $(N+1)^2$ decorrelated pink noise signals are played simultaneously over each of the $(N+1)^2$ coaxial loudspeakers of the *periphonic loudspeaker array*.

c) The playback level is adjusted such that the *LAeq*, measured over a 30s time window at the geometric center of the *periphonic loudspeaker array*, is equal to 78dBSPL(A) ± 0.5dB.

d) The reference sound pressure at the geometric center of the *free-field volume*, *p(t)*, is captured with the diffuse-field or multi-field microphone.

e) The magnitude spectrum of the reference sound pressure, *P(f)*, is calculated for the 1/12th octave intervals as given by the R40 series of preferred numbers in [6].

NOTE 2: For ideal (calibrated) loudspeakers, the *P(f)* spectra should have equal energy in each 1/12th octave intervals.

Estimated Spectrum measurement

a) The scene-based audio capture device under test is mounted in the *free-field volume* such that its geometric center coincides with the geometric center of *free-field volume* and the geometric center of the *periphonic loudspeaker array*.

b) $(N+1)^2$ decorrelated pink noise signals are played simultaneously over each of the $(N+1)^2$ coaxial loudspeakers of the *periphonic loudspeaker array*. The pink noise signals shall be identical to the signals used for the reference spectrum measurement.

c) The B-format scene-based audio format representation (compressed or uncompressed, depending on the use case being tested) is stored for offline analysis.

d) The B-format scene-based audio format representation is uncompressed (if necessary) and converted to an *equivalent spatial domain representation* of order N_{DUT} (B-Format to ESD conversion in Figure 1), where N_{DUT} corresponds to the Ambisonics order of the device under test.

e) $\hat{p}(t)$, the estimate of the sound field at the geometric center of *the free-field volume* and *periphonic loudspeaker array*, is synthesized using the *equivalent spatial domain representation* of order N_{DUT}.

NOTE 3: $\hat{p}(t)$ can be taken from the W component of the B-Format signal, as an alternative to implementing the B-Format to ESD conversion in step d).

f) The magnitude spectrum of the estimated sound pressure, $\hat{P}(f)$, is calculated for the 1/12th octave intervals as given by the R40 series of preferred numbers in [6].

Calculation of send frequency response for scene-based audio

The send frequency response for scene-based audio, *G(f)*, is calculated as $G(f) = \frac{\hat{P}(f)}{P(f)}$.

iTeh STANDARD PREVIEW (standards.iteh.ai) Full standard: https://standards.iteh.ai/catalog/standards/sist/59af9dde-97a1-4533-b74c-0fcb1c2e78fb/etsi-ts-126-260-v15.2.0-2020-03

4.1.1.4 Test method with loudspeaker array and turn table

4.1.1.4.1 Test Conditions

Loudspeaker array

a) A calibrated *loudspeaker array* shall be placed within the *free-field volume*.

b) The *loudspeaker array* shall comprise one or several semi-arcs having a radius greater or equal than 1 meter. The radius shall be reported.

c) The *loudspeaker array* shall be composed of *N+1* loudspeaker elements. The ambisonic order *N* shall be reported.

d) Each loudspeaker in the array shall be calibrated with a frequency response of [at least 100 Hz-20,000 Hz] and minimum phase response.

e) The coordinates of the loudspeaker elements are defined according to a Gaussian spherical grid [7] of order *N*. Directions shall comply with Annex B.1 and the *N+1* elevations of the spherical grid shall be reported.

Turn table

a) A turn table with a resolution of 0.5 degrees shall be used. The rotation axis of the turn table and the vertical axis of the semi-arcs shall be aligned The turn table shall be adjusted in height so that the device under test is positioned at the geometric center of the *loudspeaker array*.

b) For measurement, an azimuth step of 180/(*N*+1) degrees shall be used.

4.1.1.4.2 Measurement