**3GPP TSG- Meeting #**

**, , -**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.3* | | | | | | | | |
| **CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  |  | **CR** |  | **rev** |  | **Current version:** |  |  |
|  | | | | | | | | |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* | | | | | | | | |
|  | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **x** | Radio Access Network |  | Core Network |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| ***Title:*** |  | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | , Qualcomm Incorporated, Nokia Corporation, Dolby Sweden AB, Orange | | | | | | | | | |
| ***Source to TSG:*** | S4 | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** |  | | | | |  | ***Date:*** | | |  |
|  |  | | | |  | |  | | |  |
| ***Category:*** |  |  | | | | | ***Release:*** | | |  |
|  | *Use one of the following categories:* ***F*** *(correction)* ***A*** *(mirror corresponding to a change in an earlier release)* ***B*** *(addition of feature),* ***C*** *(functional modification of feature)* ***D*** *(editorial modification)*  Detailed explanations of the above categories can be found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | | | | | | | | *Use one of the following releases: Rel-8 (Release 8) Rel-9 (Release 9) Rel-10 (Release 10) Rel-11 (Release 11) … Rel-17 (Release 17) Rel-18 (Release 18) Rel-19 (Release 19)  Rel-20 (Release 20)* | |
|  |  | | | | | | | | | |
| ***Reason for change:*** | | Test methods for IVAS-enabled UEs are necessary to ensure proper acoustical and electrial performance/quality of terminals. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | | Test methods and corresponding test definitions for IVAS-enabled UEs are provided. In addition, several definitions regarding test equipment, environment, arrangement and signals are included. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | | Test methods for IVAS-enabled UEs are not available in Release 18. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | | 1, 2, 3, 4.0 (new), 4.1.1.3, 5 (new), Annex C (new) | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **Y** | **N** |  | | | |  | | |
| ***Other specs*** | |  | **x** | Other core specifications | | | | TS/TR ... CR ... | | |
| ***affected:*** | |  | **x** | Test specifications | | | | TS/TR ... CR ... | | |
| ***(show related CRs)*** | |  | **x** | O&M Specifications | | | | TS/TR ... CR ... | | |
|  | |  | | | | | | | | |
| ***Other comments:*** | | A new accompanying TS 26.261 on performance requirements is jointly provided for Release 18. | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | |  | | | | | | | | |

# 1 Scope

The present document specifies objective test methodologies for 3GPP immersive audio systems including channel based, object based, scene-based, parametric and hybrids of these formats. The objective evaluation methods described in the present document are applicable to audio capture, coding, transmission and rendering as indicated in their corresponding clauses. They also include testing of IVAS-based UEs [26].

# 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/R 1996-1972: "Acoustics – Assessment of noise with respect to community response".

[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

[8] M. Poletti, “Unified Description of Ambisonics Using Real and Complex Spherical Harmonics,” Ambisonics Symposium 2009, June 25-27, 2009, Graz, Austria.

[9] Recommendation ITU-T G.100.1 (06/2015): "The use of the decibel and of relative levels in speechband telecommunications".

[9] Recommendation ITU-T P.56 (12/2011): "Objective measurement of active speech level".

[10] Recommendation ITU-T P.57 (06/2021): "Artificial ears".

[11] Recommendation ITU-T P.58 (03/2023): "Head and torso simulator for telephonometry".

[12] Recommendation ITU-T P.79 (11/2007): "Calculation of loudness ratings for telephone sets".

[13] Recommendation ITU-T P.501 (05/2020): "Test signals for use in telephonometry".

[14] Recommendation ITU-T P.581 (07/2022): "Use of head and torso simulator for hands-free and handset terminal testing".

[15] Recommendation ITU-T P.340 (05/2000): "Transmission characteristics and speech quality parameters of hands-free terminals".

[16] Recommendation ITU-T P.341 (03/2011): "Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals".

[17] Recommendation ITU-T P.380 (07/2022): "Electro-acoustic measurements on headsets".

[18] Recommendation ITU-T P.381 (03/2023): "Technical requirements and test methods for analogue wired headsets or headphones and corresponding universal interface of terminals".

[19] Recommendation ITU-T P.382 (03/2023): "Technical requirements and test methods for analogue wired multi-microphone headsets or headphones and corresponding universal interface of terminals".

[20] Recommendation ITU-T P.383 (03/2023): "Technical requirements and test methods for digital headsets or headphones and corresponding interfaces of terminals".

[21] Recommendation ITU-T P.700 (06/2021): "Calculation of loudness for speech communication".

[22] Recommendation ITU-R BS.1770-5 (11/2023): "Algorithms to measure audio programme loudness and true-peak audio level".

[23] IEC 60268-1:1985: "Sound system equipment. Part 1: General".

[24] 3GPP TS 26.131: "Speech and video telephony terminal acoustic test specification".

[25] 3GPP TS 26.132: "Speech and video telephony terminal acoustic test specification".

[26] 3GPP TS 26.250: "Codec for Immersive Voice and Audio Services - General overview".

[27] 3GPP TS 26.253: " Codec for Immersive Voice and Audio Services; Detailed Algorithmic Description incl. RTP payload format and SDP parameter definitions".

[28] 3GPP TS 26.254: "Codec for Immersive Voice and Audio Services - Rendering".

[29] 3GPP TS 26.258: "Codec for Immersive Voice and Audio Services; C code (floating-point)".

[30] ETSI TS 103 224: "A sound field reproduction method for terminal testing including a background noise database".

[31] USB Implementors' Forum: "HID Usage Tables for USB", Version 1.5.

# 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 from the origin, the azimuth in mathematical positive orientation (counter-clockwise) and the elevation angle 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.

**Coded Formats**: See payload format in 3GPP TS 26.253 [27] Annex A.

**User Capture**: The user's voice is intended to be captured by the UE. The sound source is always a HATS equipped with an artificial mouth and typically only speech (or speech-like) test signals are used.

**Spatial Capture**: Acoustic scenes that include directional and/or diffuse sound sources intended to be captured by the UE.

## 3.2 Symbols

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

(A) A-weighting

dB Decibel

dBm Electrical level in dB, referenced to 1 mW

dBm0 Digital overload point in dB

dBr Electrical level at POI in dB, relative to overload point

dBov Digital level in dB, relative to overload point

dBPa Sound pressure level in dB, referenced to 1 Pa

dBSPL Sound pressure level in dB, referenced to 2E-5 Pa

*f* Frequency (in Hertz)

Hz Unit of frequency (Hertz)

Frequency response of measured versus reference signal

LAeq the sound level in dB equivalent to the total A-weighted sound energy measured over a stated period of time.

Leq the sound level in dB equivalent to the unweighted sound energy measured over a stated period of time.

Ω Electrical resistance in Ohm

Spectral magnitude of measured signal

Spectral magnitude of reference signal

Spectral magnitude of measured signal (k-th channel)

Pa Unit of pressure (Pascal)

ϕ azimuth angle (phi)

*r* distance from a point in space to the origin of the spherical coordinate system (radius of sphere)

RCV Receiving (direction)

SND Sending (direction)

θ elevation angle (theta)

TR UE delay in receiving direction

TTER Test equipment delay in receiving direction

TS UE delay in sending direction

TTES Test equipment delay in sending direction

## 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].

EXT External output

FOA First-Order Ambisonics

HATS Head-And-Torso Simulator

HFRP Hands-free Reference Point

HID Human Interface Device

HOA2 Higher-Order Ambisonics (2nd order)

HOA3 Higher-Order Ambisonics (3rd order)

HRP HATS Reference Point

ISM Independent Stream with Metadata

IVAS Immersive Voice and Audio Services

ISAR Immersive Audio for Split Rendering Scenarios

LKFS Loudness, K-weighted, relative to Full Scale

MASA Metadata-Assisted Spatial Audio

MRP Mouth Reference Point

N3D full three-dimensional normalization (of an Ambisonics signal)

NR Noise Rating

OMASA Metadata-assisted spatial audio with ISM

OSBA Scene-based audio with ISM

SBA Scene-Based Audio (Ambisonics)

SN3D Schmidt semi-normalization (of an Ambisonics signal)

SS System Simulator

TC Transport Channel (for MASA)

THD Total Harmonic Distortion

USB Universal Serial Bus

# 4 Objective Test Methodologies for Immersive Audio Systems

## 4.0 General

### 4.0.1 Applicability

This clause describes general objective test methodologies for immersive audio systems. For testing IVAS-based devices, refer to clause 5.

### 4.0.2 Test equipment

[clause 3.3 from PDoc + further specifications]

Unless specified otherwise, the accuracy of electric and acoustic measurements made by test equipment shall meet the requirements defined in clause 5.3 of 3GPP TS 26.132 [25].

For tests with head tracking, HATS rotation around the vertical axis should be realized using a motorized turntable or a HATS with motorized head rotation. For motorized or manual rotations of HATS and/or UE, error in orientation (elevation and azimuth) shall not exceed [±2°]

NOTE: A motorized rotation of HATS and/or UE is highly recommended. Some UE may not have a natural reference orientation (which, for instance, may be defined by the direction of a screen). In this case, the UE may reset the reference direction automatically, e.g., to the primary device orientation over a previous span of time. This has to be taken care of during the measurement. The measurement with the rotated HATS and/or UE has to be performed quickly enough to prevent the reference direction from being spuriously readjusted and benefits from automation.

Head-and-torso simulators (HATS) used for acoustic testing are specified in ITU-T Recommendation P.58 [11], corresponding artificial ears for testing in receiving direction are specified in ITU-T Recommendation P.57 [10] (Type 3.3 or Type 4).

In sending direction, HATS equipped with mouth simulators (or equivalent stand-alone mouth simulator) used as a single sound source for spatial or user capture (see clause 5.4.1) in the test arrangement shall comply with ITU-T P.58 [11] and ITU‑T P.581 [14].

Loudspeakers used as a single sound source in the test arrangement for spatial capture (see clause 5.4.1) shall meet the following requirements:

- The spectrum of the acoustic signal produced by the loudspeaker shall be equalized under free field conditions with a measurement microphone positioned on the main loudspeaker axis at a distance of 50 cm from the loudspeaker membrane. The achieved free field equalized spectrum in 1/3rd octave bands shall be within ±1 dB from 100 Hz to 200 Hz and shall be within ±0.5 dB from 200 Hz to 20 kHz.

- THD ≤ [1]% when measured at 1 m on axis with an [85] dBSPL sinusoidal signal for frequencies ≥ 125 Hz.

- Maximum long term level ≥ [96] dBSPL when measured at 1 m on axis with the simulated programme signal defined in clause 7 of IEC 60268-1 [23].

### 4.0.3 Environmental conditions

The test environment (anechoic chamber) 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. Qualification or verification of this requirement can be conducted using the methods and limits for deviation from ideal free-field conditions as specified in ISO 3745 [3] or ITU-T P.340 [15].

For testing headset UE in receiving direction, the test environment according to clause 6.1 of ETSI TS 103 224 [30] may be used.

NOTE: The usage of the test environment according to ETSI TS 103 224 for other UE types and use cases is for further study.

Following the noise rating (NR) determination procedures in [4], The equivalent continuous sound level of the test environment in each 1/3rd octave band, *Leq(f)*, shall be less than the limits of the NR15 curve and should be less than the limits of the NR10 curve.

#### 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* shall be equal to 4.

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.

5 Objective Test Methodologies for IVAS-based UEs

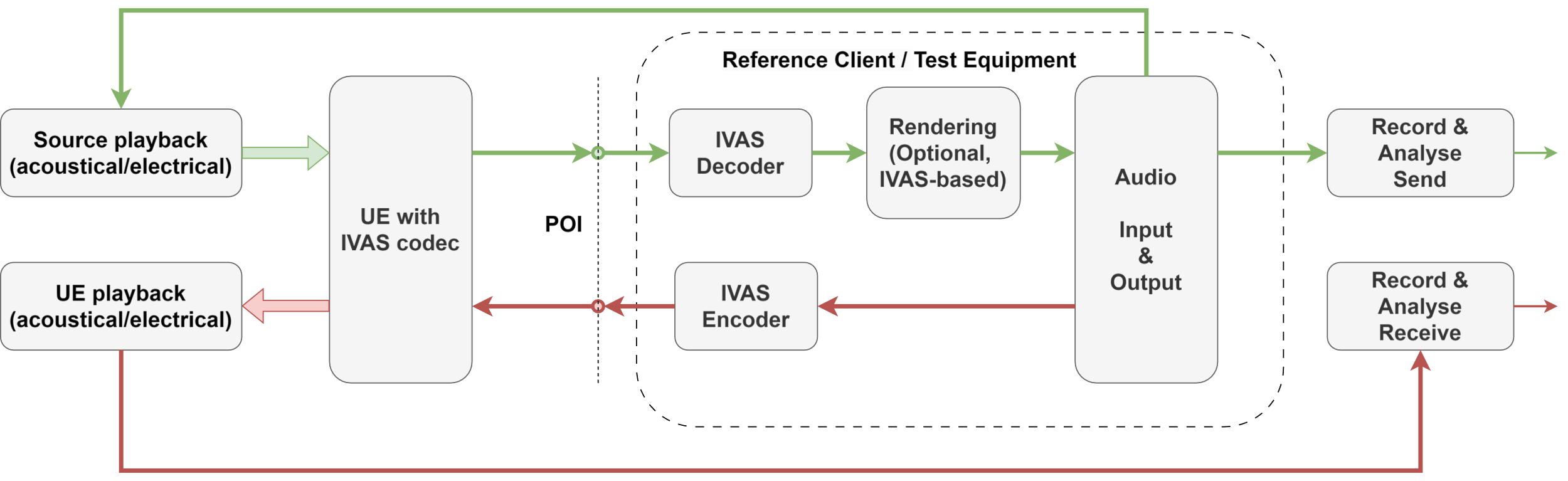
5.1 Overview

This clause adds objective test methodologies that are specific to IVAS-based UEs. The following clauses allow testing of UEs on either an acoustical or an electrical interface. In case of acoustical interface testing, a HATS (head and torso simulator) is used for a realistic simulation of an average user. A motorized turntable or an HATS with motorized head rotation is used to test head-tracking behavior of the UE under test. Moreover, in contrast to conventional telephony, which aims to capture/present the subscriber's voice isolated from any acoustical background scene, IVAS-based UEs aim to capture and/or reproduce spatial information of the acoustic scene.

NOTE: The tests in this clause are limited to single sound source tests and are not intended to evaluate UE performance in more complex sound fields. Test methods with more complex sound fields are for further study.

5.2 Interface Definitions

UE testing is realized by connecting a terminal to a test system composed of a system simulator and a reference client. The system simulator simulates the access network, provides core network functionalities and a point of interconnection (POI). The reference client serves as the far-end communication endpoint at the POI and provides IVAS encoder, decoder and rendering functionalities. Test sequences are both captured and fed into the reference client for sending and receiving direction tests, as illustrated in Figure 6. Immersive audio encoding, decoding, and rendering can be part of the system simulator or the reference client. No further transcoding beyond linear PCM shall take place.



**Figure 6: General test setup with UE and reference client**

The digital overload point of 3.14 dBm0 (corresponds to 0 dBr according to ITU-T G.100.1 [9]) for testing IVAS-based UEs applies equally for all possible encoded or decoded signal channels. The same definition as in 3GPP TS 26.132 [25] is used:

D/A converter - a digital test sequence representing the codec equivalent of an analogue sinusoidal signal with an RMS value of 3.14 dB below the maximum full‑load capacity of the codec shall generate 0 dBm across a 600 Ω load;

A/D converter - a 0 dBm signal generated from a 600 Ω source shall give the digital test sequence representing the codec equivalent of an analogue sinusoidal signal with an RMS value of 3.14 dB below the maximum full‑load capacity of the codec.

The details of the acoustical and electrical test setup as well as IVAS session parameters are UE-type-dependent and are given in the following clauses.

5.3 Test conditions

5.3.1 Environmental conditions

The same requirements as specified in clause 4.0.3 apply for IVAS-based testing.

5.3.2 System simulator and reference client

The system simulator configuration and radio conditions on the air interface shall be as specified in clause 4.2 of 3GPP TS 26.132 [25]. Unless otherwise specified for the respective test, the system simulator shall provide an error-free radio connection to the UE under test.

The UE shall be connected to a system simulator and test equipment supporting the IVAS codec [26]. Unless specified otherwise, 48 kHz sampling rate and fullband mode shall be used.

Since UEs may provide different capabilities for sending (capture) and receiving (rendering) direction, a bidirectional communication may not be symmetric regarding IVAS Coded Formats and corresponding bitrates.

During negotiating and exchange of session parameters via Session Description Protocol (SDP), the UE advertises its supported and preferred Coded Formats in sending and receiving direction, as well as corresponding bitrate ranges.

- The reference client shall advertise via SDP the envisioned IVAS audio format for testing sending and receiving direction along with corresponding bitrate as specified in Table 1 as the preferred and only capabilities.

- For sending direction, the decoder in the reference client shall be configured for the EXT decoder output format.

- For receiving direction, the encoder in the reference client shall be configured for the Coded Format envisioned for testing.

**Table 1: Bitrates per audio format used for testing**

|  |  |  |  |
| --- | --- | --- | --- |
| **Coded Format** | **Subformat** | **Default Bitrate for testing…** | **Max. bitrate [kbit/s]** |
| Stereo | Stereo | TBD | 256 |
|  | Binaural | TBD | 256 |
| ISM | 1 | TBD | 128 |
| 2 | TBD | 256 |
| 3 | TBD | 384 |
| 4 | TBD | 512 |
| SBA | FOA | TBD | 512 |
| HOA2 | TBD | 512 |
| HOA3 | TBD | 512 |
| MASA | 1 TC | TBD | 512 |
| 2 TC | TBD | 512 |
| OSBA |  | TBD | 512 |
| OMASA |  | TBD | 512 |
| Multichannel | 5.1 | TBD | 512 |
| 7.1 | TBD | 512 |
| 5.1.2 | TBD | 512 |
| 5.1.4 | TBD | 512 |
| 7.1.4 | TBD | 512 |
| NOTE: The maximum bitrates are provided for information. | | | |

NOTE 1: Default bitrate configurations for testing are for further study.

NOTE 2: The maximum bitrates listed in Table X may not be supported by all UEs.

NOTE 3: The Coded Format ISAR (see [26]) is for further study.

The bitrate shall be provided by the network operator configuration. If the bitrate is not available from the operator, the maximum bitrate supported by the UE shall be used. The test operator shall report the bitrate used for testing.5.3.3 Test equipment

The same requirements as specified in clause 4.0.2 apply for IVAS-based testing.

5.4 Test arrangement

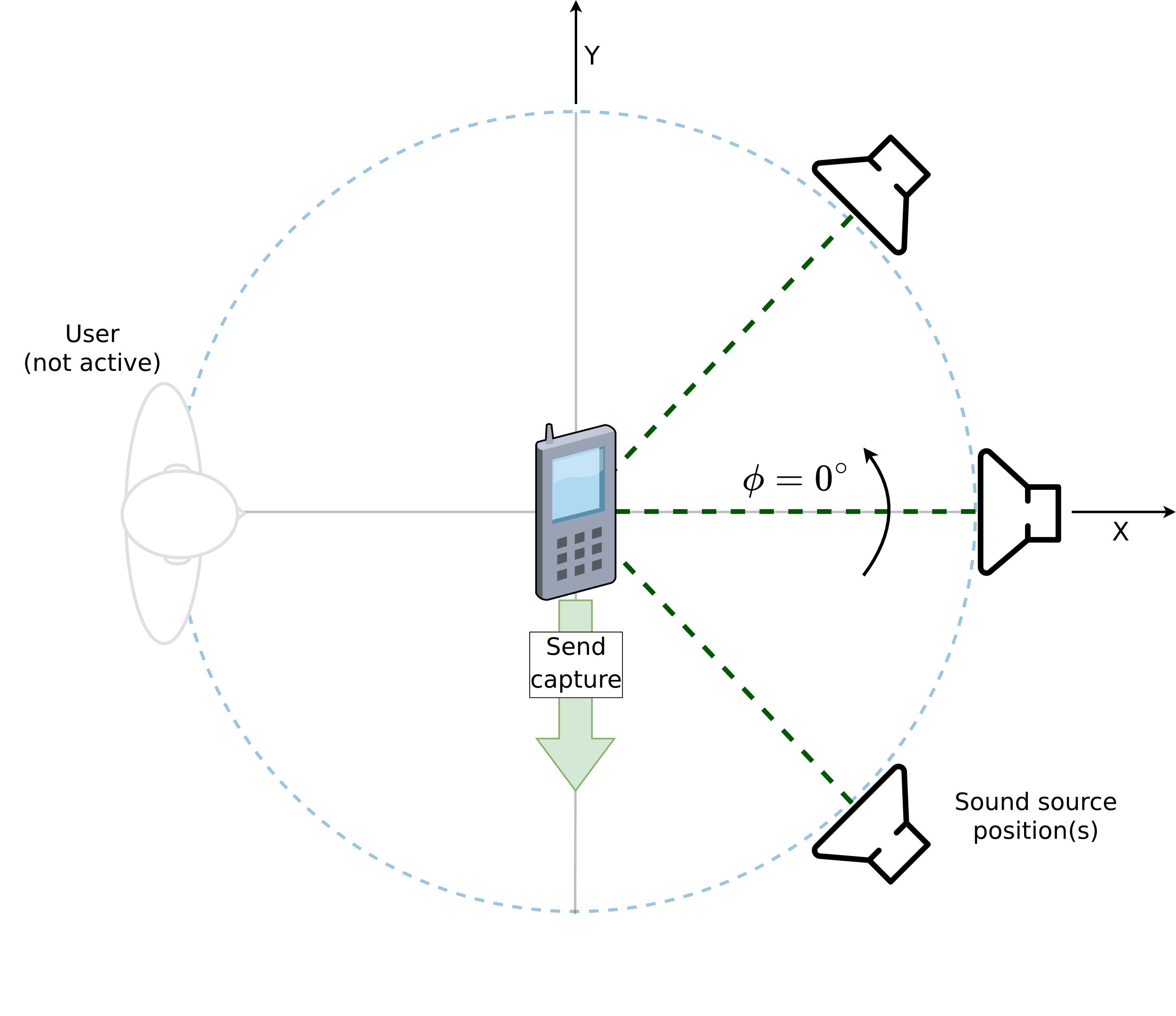
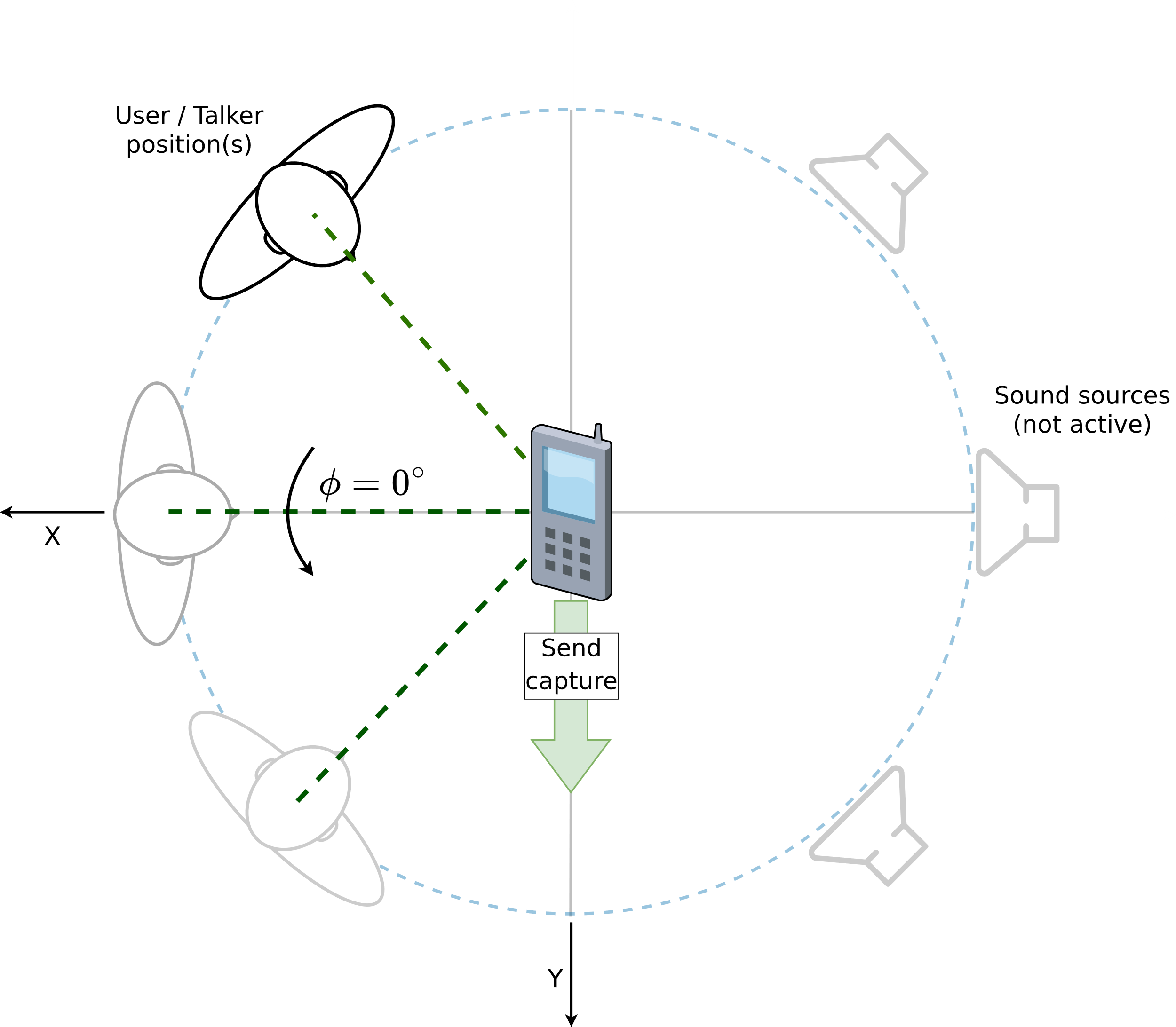
5.4.1 Capture modes

To simulate a single sound source, either loudspeaker or a HATS equipped with an artificial mouth (or equivalent stand-alone mouth simulator) positioned relative to the UE at a certain anlge/distance. In general, the default source direction (independent of UE type or capture mode) is located at ϕ = 0° and θ = 0°. Two different types of sound sources are distinguished:

1) User capture: The user's voice is intended to be captured by the UE. The sound source is always a HATS equipped with an artificial mouth and typically only speech (or speech-like) test signals are used.

2) Spatial capture: Acoustic scenes that include directional sound sources and/or diffuse sound field, are intended to be captured by the UE.

The general principle of both capture modes is illustrated in Figure 7.

****

**Figure 7: Device orientation and sound sources for User Capture (left) and Spatial Capture (right)**

NOTE: Figure 7 is indicative and the illustrated example does not refer to any specific test arrangement or method. Diffuse sound fields for Spatial Capture are out of scope for IVAS-based UE testing in the present document.

5.4.2 UE types and positioning

5.4.2.1 Overview

The classification of an UE type is defined as follows:

- An UE type is composed of *SND-UE-type* and *RCV-UE-type*.

- The SND-UE-type is defined as the combination of a certain audio capturing mode (acoustic or electric) and a negotiated coded format (see clause 5.3.2).

- The RCV-UE-type is defined as the combination of a negotiated coded format (see clause 5.3.2) and a certain audio playback mode (acoustic or electric).

- Each audio capturing/playback mode corresponds to a specific physical test arrangement.

NOTE 1: The definition of IVAS-enabled UE types is more complex than in 3GPP TS 26.132 [25]. Due to the variety of new applications that are enabled by the IVAS codec, the capture and playback audio format may not necessarily be the same in sending and receiving direction, which explains the classification introduced above.

UEs may support multiple IVAS formats in sending and receiving direction, which are negotiated during call setup. At least one supported IVAS format shall be tested in all directions supported by the UE, which is selected according to the following priority:

1) Format specified by the manufacturer (if applicable).

2) Preference of the UE, as indicated during negotiation in SDP.

3) Test operator selects format based on form factor and envisioned use case of the UE.

The IVAS format for each tested direction shall be documented in the test report. Other available supported formats may be tested as well to ensure best-possible compatibility with other UE types.

Complementary to the well-defined IVAS formats, capturing/playback modes and corresponding interfaces are given in the following clauses along with several UE type definitions, which may be applicable in SND and/or RCV. All UE type definitions with acoustical interfaces assume that microphones and loudspeakers/headset of the UE are either integrated into the device or that necessary additional off-the-shelf equipment (like e.g., headset, microphone array, loudspeaker array) is either bundled with the device or explicitly recommended by the manufacturer.

If no acoustical interface is available, the electrical interface shall be tested and the test setup according to clause 5.4.2.7 applies.

NOTE 2: If testing at the electrical interface is not possible for important technical reasons (e.g., due to a non-standardized electrical interface), an acoustic test may be carried out with suitable third-party equipment, which is, e.g., recommended by the manufacturer. This allows to perform tests at an acoustic interface in order to informally assess the performance of the UE (without applying requirements).

The physical test arrangement used for UE testing in sending and receiving direction is in general specified by the manufacturer by:

- Referencing one of the following clauses (recommended),

- Referencing a test arrangement from other standards (e.g., ITU-T P.340 [15] or P.341 [16]),

- Specifying an individual test arrangement.

In case no instructions on the test arrangement are provided by the manufacturer, the test operator shall select one based on the envisioned use case, form factor, etc. from either one of the following clauses or from other standards. If no suitable test arrangement can be identified for certain UEs with acoustical interface, the test operator may set up an individual arrangement or modify an existing one. In any case, the arrangement used for testing shall be documented in the test report.

5.4.2.2 Handset Mode

The EVS-compatible mono mode of IVAS shall be tested according to clause 9 of TS 26.132 [10]. Requirements according to clause 7 of 3GPP TS 26.131 [24] apply.

Immersive testing for handset UE is not considered in the present document due to the following reasons:

- RCV: Monaural listening cannot provide any spatial/immersive audio.

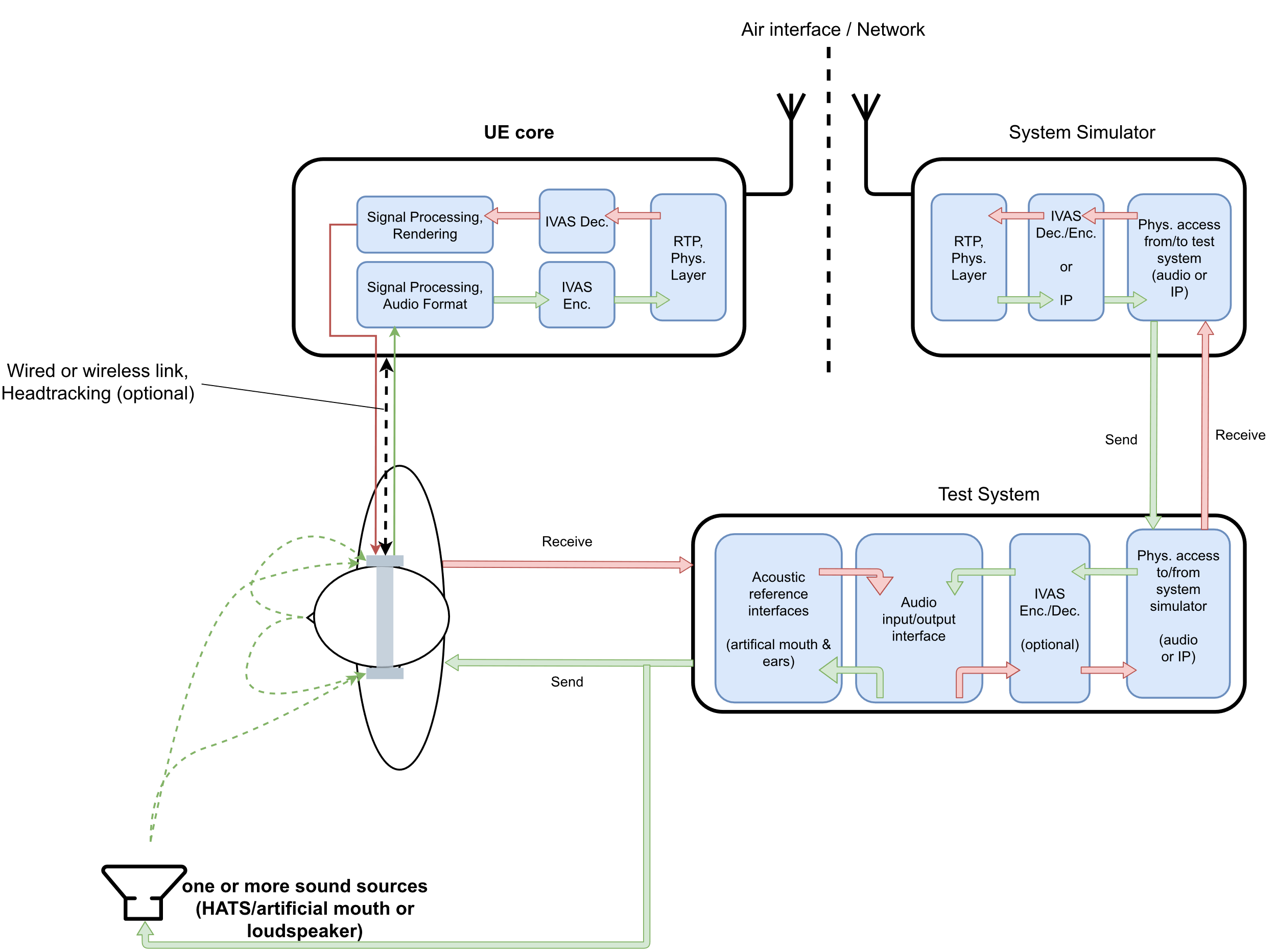
- SND: The device is typically located close to the user's head, which can limit the capture of spatial information.

NOTE: There might be some applications also in handset mode for some immersive audio formats, for example to capture ambient sound at the near end. Such scenarios are a kind of extension to the traditional mono telephony and are not excluded in general. However, so far, test methods are not specified for handset mode and are for further study.

5.4.2.3 Headset Mode

The test setup for headset UE for sending and receiving directions is shown in Figure 8. It applies to all devices that provide a head-worn acoustical frontend. The acoustical frontend may either be internal or external to the UE device. In the latter case, it may be connected via wired or wireless link (e.g., analogue jack, Bluetooth, or USB).

The device may provide head tracking data that can be used for rendering binaural audio in receiving direction. The head-worn acoustical frontend may provide headtracking data through a wired or wireless link (see Figure 8), which can be used for rendering binaural audio in receiving direction.

****

**Figure 8: Headset UE and test equipment**

**A diagram of a speaker

Description automatically generated**

**Figure 9: Sound source positioning for Headset UE tests**

**User capture**

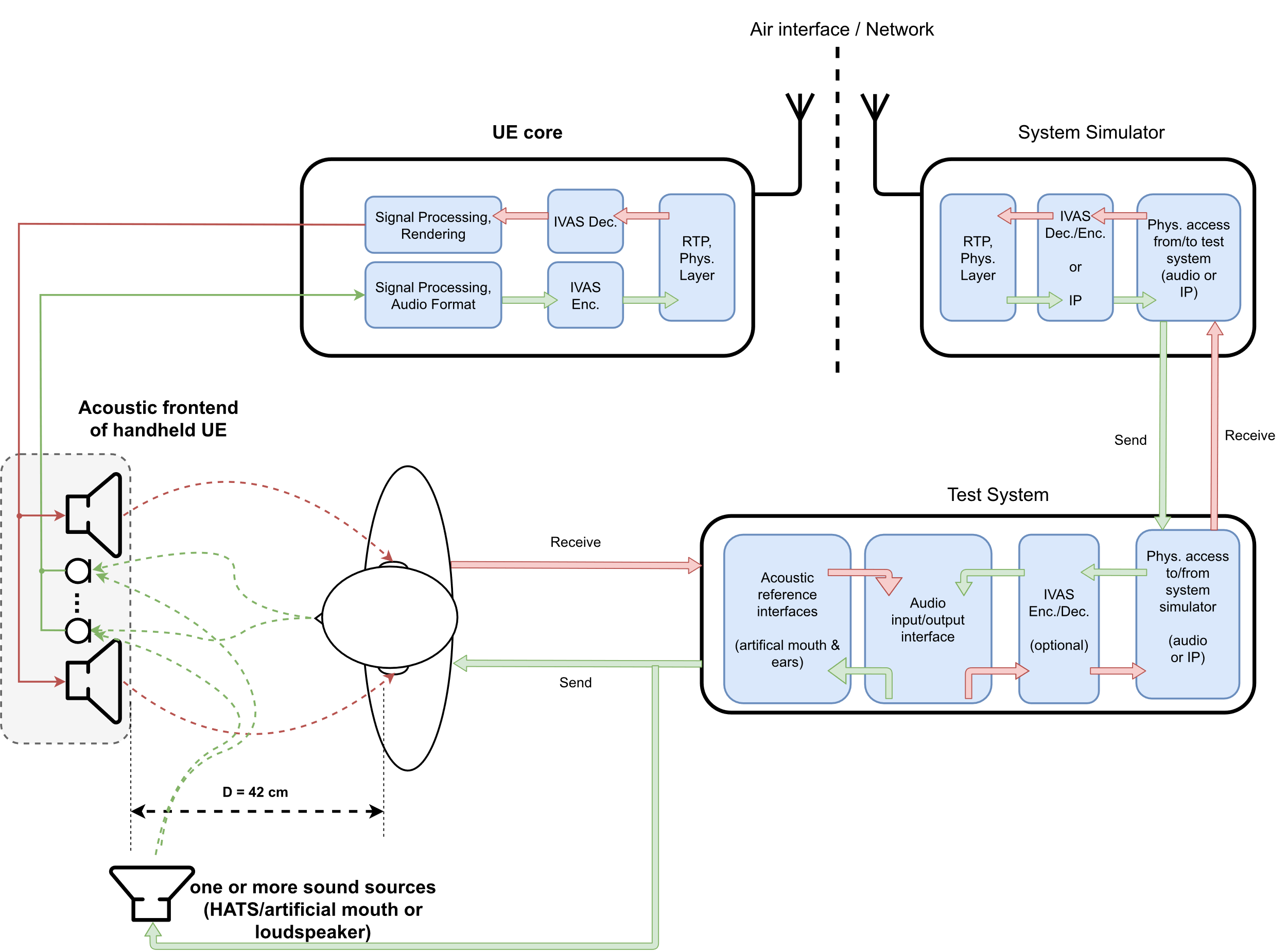
If not specified otherwise, the arrangement for a single sound source is a HATS wearing the headset. The playback level at MRP shall be calibrated to -4.7 dB Pa.

**Spatial capture**

If not specified otherwise, the arrangement for a single sound source is a loudspeaker positioned at 0° azimuth and elevation and at a distance of 1 m relative to HRP of the HATS wearing the headset device, which is assumed to be the geometric center for all headset UEs. Sound source positioning for sending tests is illustrated in Figure 9. The playback levels at HRP (in the absence of HATS) shall be calibrated to [75 dBSPL].

5.4.2.4 Handheld hands-free Mode

The test setup for handheld hands-free UE for sending and receiving directions is shown in Figure 10. It applies to all devices that can be held in front of the user.

****

**Figure 10: Handheld hands-free UE and test equipment**

**A diagram of a person's head

Description automatically generated**

**Figure 11: Sound source positioning for Handheld UE tests**

The UE orientation (landscape/portrait and top/bottom position, front/back side facing the user) used for testing is specified by the manufacturer. If such information not available, the orientation depends on the capture mode.

The test fixture (e.g., microphone stand) used to mount the handheld hands-free UE for the measurements should be as acoustically transparent as possible and should not obstruct any of the input microphones.

**User capture**

If not specified otherwise, the arrangement for a single sound source is a HATS positioned at a distance of 42 cm from the center point of the visual display of the UE (same setup as in TS 26.132 [25] for handheld hands-free). If applicable, different geometries of this setup are considered in corresponding test methods (for e.g., multi-talker scenarios or speech from certain angles). The playback levels at MRP shall be calibrated to -1.7 dB Pa.

If no manufacturer-defined orientations are defined, the UE shall be positioned in portrait mode, whereas the front side is facing the user.

**Spatial capture**

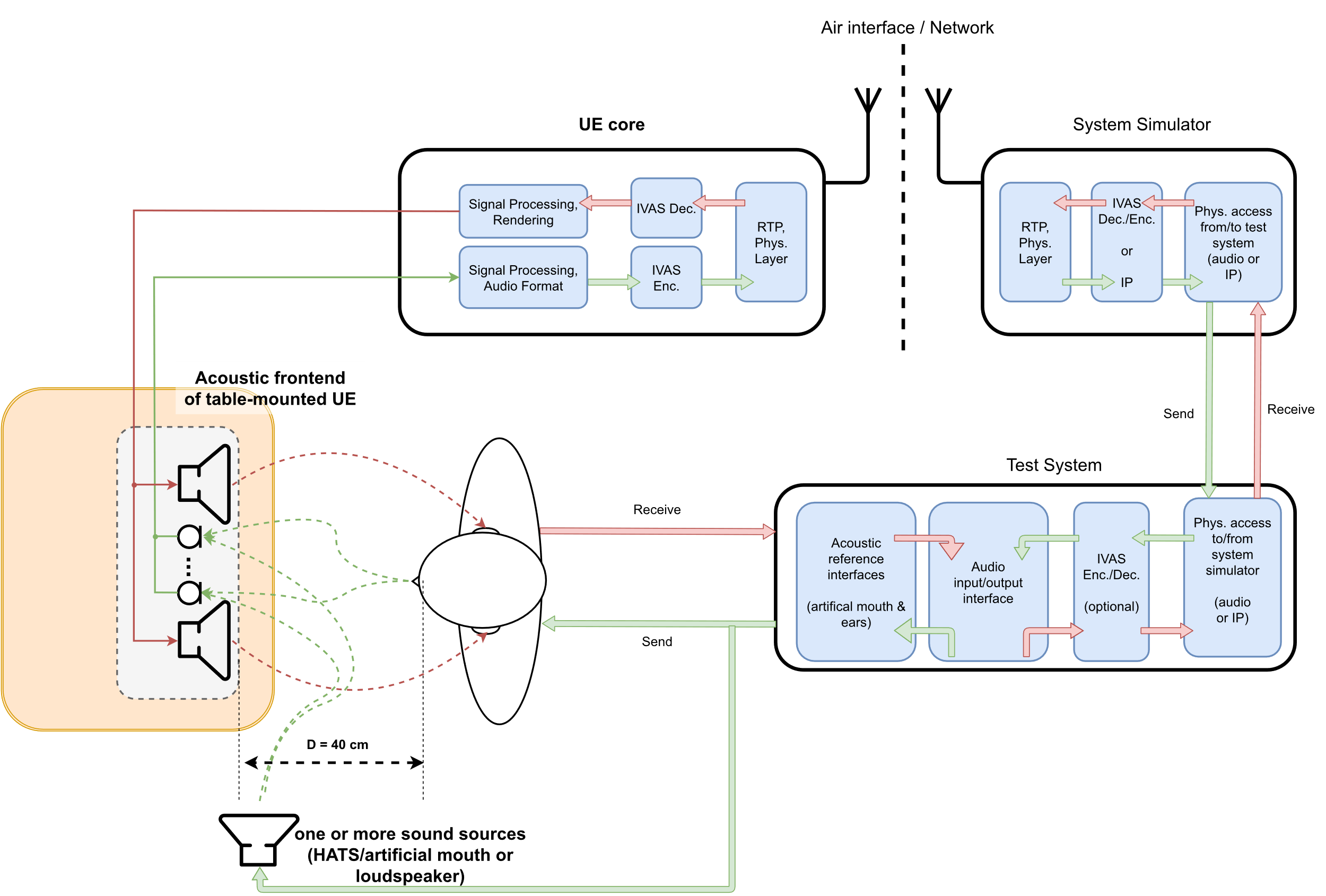
Sound source positioning for sending tests with spatial capture is shown in Figure 11. If not specified otherwise, the arrangement for a single sound source is a loudspeaker positioned at 0° azimuth and 0° elevation and at a distance of 1 m relative to the geometric center of the handheld hands-free UE. The playback levels at the UE shall be calibrated to [75 dBSPL]. If applicable, different geometries of this setup are considered in corresponding test methods (for e.g., multiple sources from different angles).

If no manufacturer-defined orientations are defined, the UE shall be positioned in landscape mode (top of the device pointing to the left), whereas the front side is facing the user.

NOTE: Spatial capture typically targets at acoustic scenes opposite to the user of the device. The acoustic impact of the simulated user/HATS in such a setup is for further study.

5.4.2.5 Table-mounted Mode

The test setup for table-mounted hands-free UE for sending and receiving direction is shown in Figure 12. It applies to all hands-free devices that are intended for usage on tables (like e.g., conference devices). In contrast to handheld hands‑free UE, the reflections of the table are explicitly included in the test setup.

****

**Figure 12: Table-mounted hands-free UE and test equipment**

Figure 12 shows an example with a distance of D = 40 cm between front of the UE and lip reference plane of the user, which corresponds to the desktop hands-free setup as specified in Recommendation ITU-T P.341 [16], which is also referenced in 3GPP TS 26.132 (width W = 40 cm, height H = 30 cm). In general, multiple sub-setups may be considered for this UE type, like e.g., the "group audio terminal" position (see clause 4.2.4 of P.341 [16]) or the softphone/laptop-based setups 3GPP TS 26.132 [25].

NOTE: The term "table-mounted hands-free" is suggested here instead of "desktop hands-free", as used in e.g., 3GPP TS 26.132. The intention for this is to explicitly address also different/larger setups like e.g., conferencing scenarios with multiple microphones and loudspeaker arrays.

**A diagram of a speaker

Description automatically generated**

**Figure 13: Sound source positioning for Table-mounted UE tests**

**User capture**

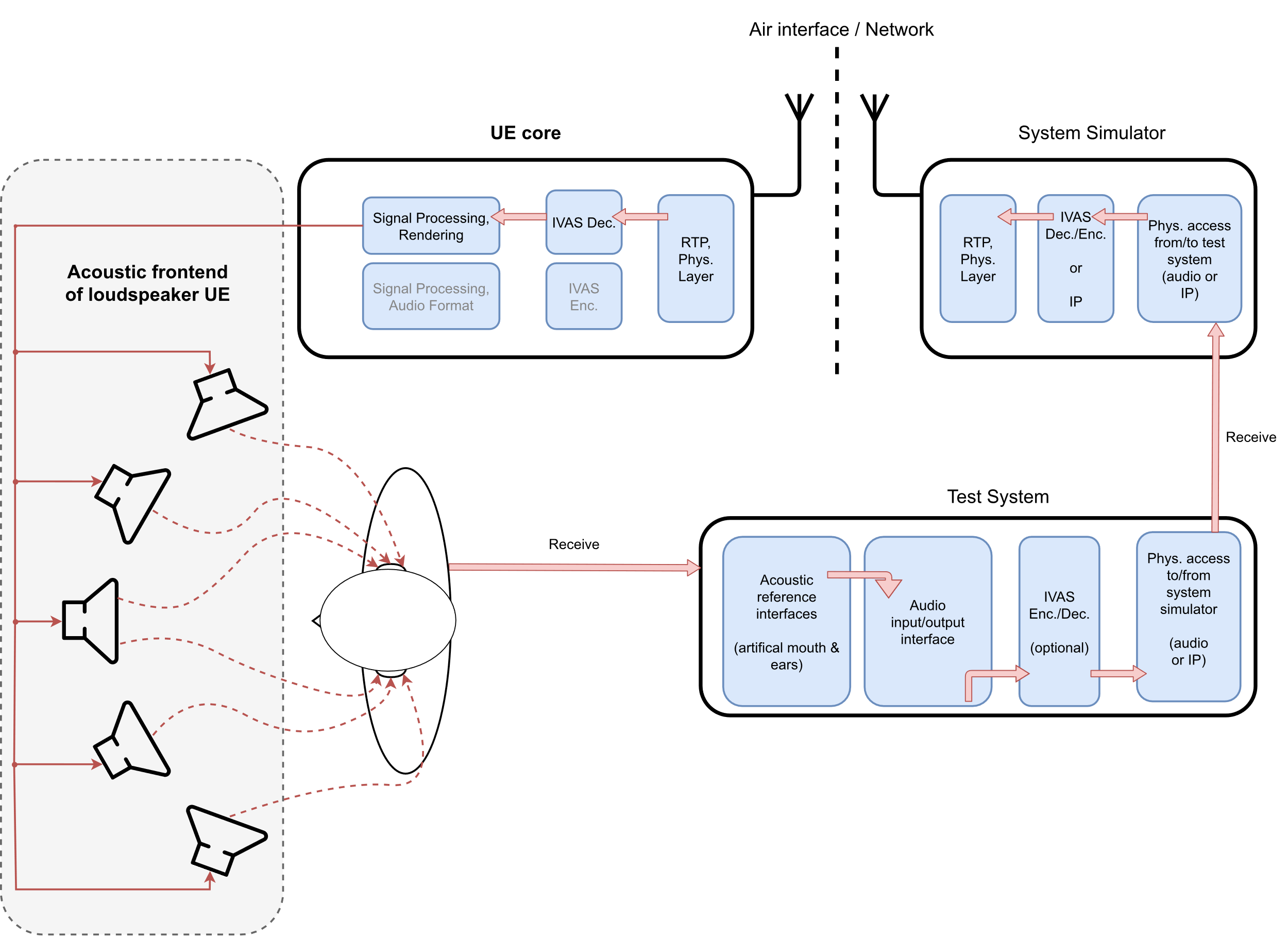
If not specified otherwise, the arrangement for a single sound source is a HATS positioned at a distance of 40 cm between UE and lip reference plane and a height of 40 cm. The playback levels at MRP shall be calibrated to ‑1.7 dB Pa.

**Spatial capture**

Sound source positioning for sending tests is shown in Figure 13. If not specified otherwise, the arrangement for a single sound source is a HATS or loudspeaker positioned at a height H = 40 cm above the table and at a width of 80 cm in front of the geometric center of the table-mounted UE. In spherical coordinate system, this corresponds to 0° azimuth, 26.6° elevation and a distance of 89.4 cm. The playback levels at the HFRP shall be calibrated to -24.7 dBPa.

5.4.2.6 Loudspeaker Mode

The test setup for loudspeaker hands-free UE for receiving direction is shown in Figure 14. It applies to multichannel loudspeaker systems and speaker arrays, e.g., soundbars or automotive infotainment systems.

****

**Figure 14: Loudspeaker hands-free UE and test equipment**

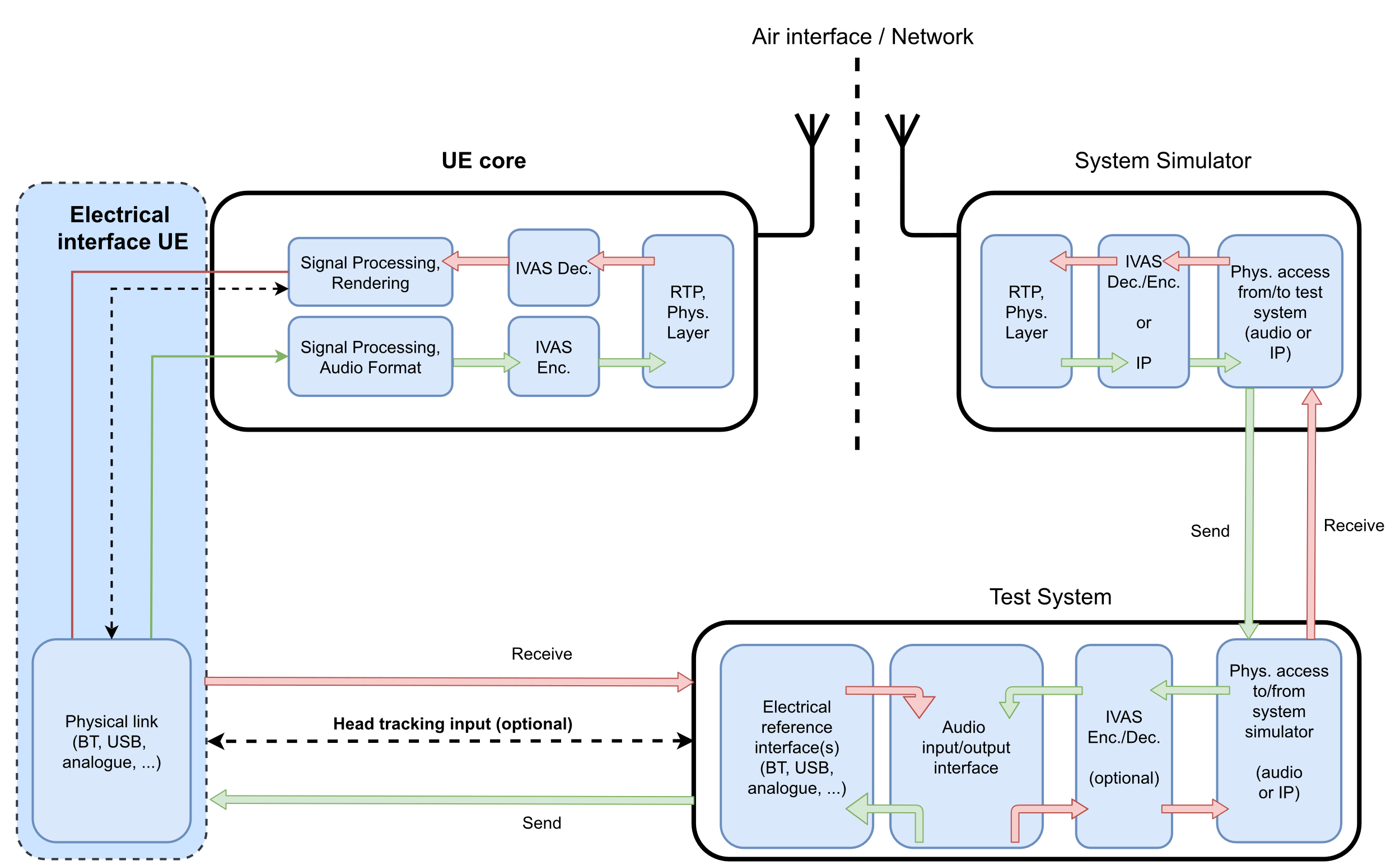
5.4.2.7 Electrical interface Mode

The test setup for electrical interface UE for sending and receiving directions is shown in Figure 15. It applies to all devices that do not provide integrated or associated equipment for capturing and/or reproduction of immersive audio. Wired or wireless digital audio interfaces according to Recommendation ITU-T P.383 [20] (e.g., Bluetooth or USB) are commonly used. Note that the interface may also be realized via an analogue jack plug, which provides up to two channels in receiving and sending direction (see Recommendations ITU-T P.381 [18] and ITU-T P.382 [19]).

The device may provide an additional input for head tracking data that can be used for rendering the receiving direction (e.g., Bluetooth or USB with HID profile [31]).

Different equipment may be connected to the electrical interface UE such that the combination of UE and additional equipment will behave like one of previous UE types (e.g., headset or loudspeaker). Test methods apply according to the envisioned use-case. The default test signal for electrical insertion of a single sound source shall correspond to the envisioned use-case.

EXAMPLE: If the electrical interface of an UE is envisioned to connect a third-party immersive headset (e.g., with headtracking functionalities), the default test signal contains a virtual single sound source at the default source position specified for headset UE (see clause 5.4.2.3). Only test methods for headset UE are applicable in this case.

****

**Figure 15: Electrical interface UE and test equipment**

5.4.3 UE configuration

For testing, the UE shall be configured for the relevant and/or envisioned use cases as described in clause 5.4.2.

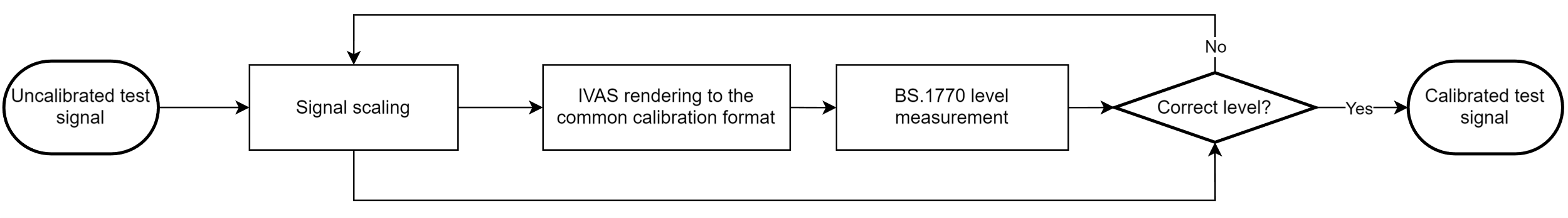
UEs shall be tested *as is,* even if they have signal enhancement features, like e.g., noise suppression, which may cause that performance requirements are not met. During the tests, any possible internal development modes of the UE bypassing these features shall be disabled. However, if such issues are encountered and the UE allows to disable certain of these features, the tests should be repeated with these features disabled to document the possible root cause of the problem.

Unless stated otherwise, if a volume control is provided in receiving direction, the setting is chosen such that the nominal receiving loudness is met as closely as possible.

5.5 Test signals

5.5.1 Test signal calibration

The input signal levels for testing receiving direction or testing with an electrical interface in sending direction shall be calibrated by means of a calibration factor, which is determined according to the iterative procedure illustrated in Figure Figure 16. The rendering of the input signal to the target format shall be done via external IVAS reference renderer, which is defined in clause 6 of 3GPP TS 26.254 [28]. The rendered output format shall be set to 7.1+4 output, except if the format of the input signal is already multichannel or stereo. The level is then calculated according to ITU‑R BS.1770 [22] for these intermediate signals. If the difference between measured and target level exceeds [0.5] LKFS the procedure is repeated for up to ten iterations.

****

**Figure 16: Test signal input level adjustment procedure for receiving tests**

For mono real speech signals, acoustical and electrical calibration of test signals shall be carried out with active speech level according to ITU-T P.56 [9], calculated over the complete test sequence.

5.5.2 Virtual positioning

For testing of a single, directional sound source in receiving direction and sending direction with an electrical interface, the test signal shall be positioned virtually to represent the tested source direction.

**Object-based audio**

Unless specified otherwise, the source position of the test signal shall be set by format specific metadata as defined in Table 2. No further metadata fields shall be set.

**Table 2: Object-based audio format specific metadata**

|  |  |  |  |
| --- | --- | --- | --- |
| **Azimuth** | **Elevation** | **Spread** | **Gain** |
| According to the tested azimuth | According to the tested elevation | 0 | 1 |

**Scene-based audio**

Unless specified otherwise, the source position of the test signal shall be set by a multi-component Ambisonics (ACN/SN3D) signal that represents a source from the particular incidence angle. To generate the test signal, first the virtually positioned object-based format is created. Then the IVAS external renderer according to 3GPP TS 26.254 [28] is used to obtain the desired signal in scene-based audio output format from the object-based input.

**Metadata-assisted spatial audio**

Unless specified otherwise, the source position of the test signal shall be set by the format specific metadata. The applied descriptive metadata for every frame and the applied spatial metadata for every time-frequency tile shall be set as defined in Table 3 and Table 4. The fields of the MASA metadata format is specified in Annex A of 3GPP TS 26.258 [29]. Unless specified otherwise, the source signal shall be applied on each transport channel used for testing.

**Table 3: Applied descriptive metadata**

|  |  |
| --- | --- |
| **MASA format descriptive common metadata parameters** | **Assigned values for every metadata frame** |
| Format descriptor | Default |
| Number of directions | 1 (bit value 0) |
| Number of channels | 1 or 2 (bit value 0 or 1), depending on the number of applied transport channels |
| Source format | Bit values 00 (Default/unknown) |
| Variable description | 12 bit zero-padding (Default/unknown) |

**Table 4: Applied spatial metadata**

|  |  |
| --- | --- |
| **MASA format spatial metadata parameters** | **Assigned values for every time-frequency tile in all MASA metadata sub-frames** |
| Direction Index | According to the tested azimuth and elevation |
| Direct-to-total energy ratio | 1.0 |
| Spread coherence | 0.0 |
| Diffuse-to-total energy ratio | 0.0 |
| Surround coherence | 0.0 |
| Remainder-to-total energy ratio | 0.0 |

**Multichannel audio**

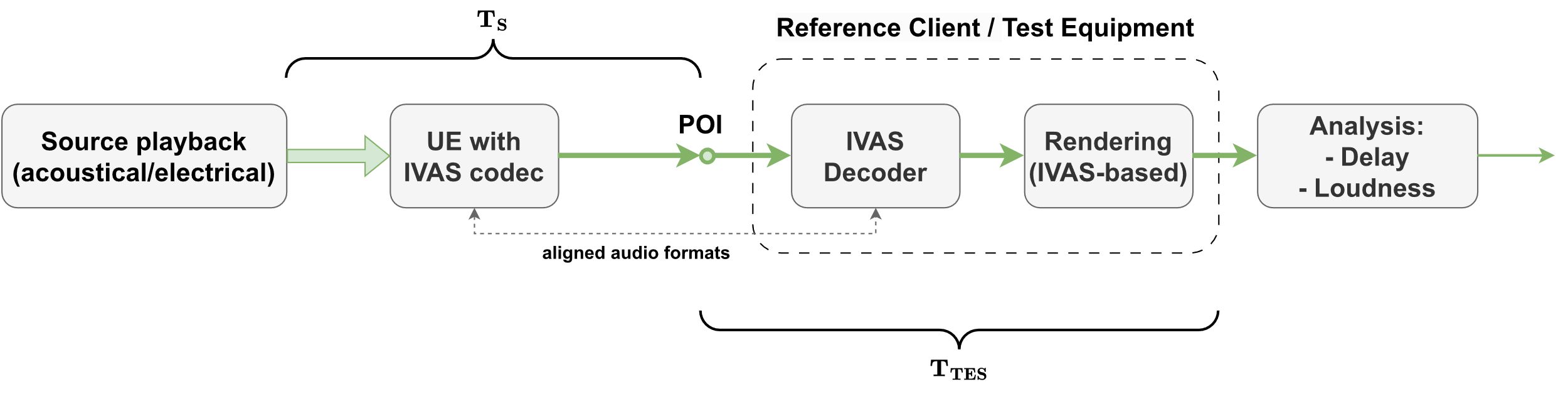
To generate the test signal, first the virtually positioned object-based format is created. Then the IVAS external renderer according to 3GPP TS 26.254 [28] is used to obtain the desired signal in multichannel audio output format from the object-based input.

5.6 Test methods for sending direction

5.6.1 Delay

The assessment of sending UE loudness is illustrated in Figure 17. The default arrangement for a single sound source is used, as defined in clause 5.4.2 for each UE type. The IVAS audio format used by the UE is first decoded in the reference client and then rendered to a common analysis-dependent format. For this rendering step, the IVAS renderer shall be used.

NOTE: The default arrangement for a single sound source only depends on the UE type but is independent of the used IVAS audio format.

****

**Figure 17: Test setup for sending loudness and delay**

**Test method**

1. The default arrangement for a single sound source as defined in clause 5.4.2 is set up according to the UE type.
2. The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [xx] and is calibrated to an active speech level according to Recommendation ITU-T P.56 [9] as defined in clause 5.4.2 for the corresponding UE type.
3. The UE under test and the reference client are connected and configured as described in the claus, except the renderer configuration shall be set to the mono.
4. The UE delay in sending direction TS is obtained between the acoustical sound source and the electrical POI of the test equipment. The source signal is used as a reference for the cross-correlation analysis described in Annex C, which is used to determine TS.
5. The measured delay shall be compensated by the delay TTES introduced by the test equipment (including possible contributions of the rendering to mono).

5.6.2 Loudness

The assessment of sending UE delay is illustrated in Figure 17. The default arrangement for a single sound source is used, as defined in clause 5.4.2 for each UE type. The IVAS audio format used by the UE is first decoded in the reference client and then rendered to a common analysis-dependent format. For this rendering step, the IVAS renderer shall be used.

**Test method**

1. The default arrangement for a single sound source as defined in clause **Error! Reference source not found.** is set up according to the UE type.
2. The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [xx] and is calibrated to an active speech level according to Recommendation ITU-T P.56 [9] as defined in clause 5.4.2 for the corresponding UE type.
3. The UE under test and the reference client are connected and configured as described in the clause [**Error! Reference source not found.**], except the renderer configuration shall be set to the mono.
4. The UE loudness in sending direction is obtained by [Analysis method TBD: BS.1770, P.700, P.79 SLR, P.56 ASL, etc.]

5.6.3 Frequency response (single source)

5.6.3.1 Test method

The default [spatial capture] arrangement for a single sound source is used, as defined in clause 5.4.2 for each UE type. The decoded and rendered output format shall be the same as the IVAS audio format used by the UE.

For each source direction under test, the following procedure shall be used:

1. The test signal to be used for the measurements shall be [the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [9], calibrated to an active speech level according to clause 5.5.1.
2. The UE under test and the reference client are connected and configured as described in the clause **Error! Reference source not found.**.
3. **Acoustical Interface:** The UE is mounted as described in the clause [**Error! Reference source not found.**] and the acoustic source is positioned such that the source direction under test is met. The test signal is played via the acoustic source.

**Electrical Interface:** The test signal is generated by virtually placing the acoustic source such that the source direction under test is met as described in clause [**Error! Reference source not found.**].

1. The frequency spectrum of the decoded output is calculated for the 1/12th octave intervals as given by the R40 series of preferred numbers in [ISO 3] for frequencies from [100 Hz] to [12 kHz] inclusive. The output format dependent frequency spectrum calculation shall be done as defined in [4.1.3].

5.6.3.2 IVAS format specific definitions

**Stereo**

[

Frequency spectrum of the stereo audio signal is defined as a ratio of the sound pressure magnitude spectrum of the audio channels (): and a reference magnitude spectrum . Letter denotes audio channel number. Thus, for each stereo audio signal channel a frequency response is determined by:

]

**Object-based audio**

[

Frequency spectrum of the Object-based audio signal is defined as a ratio of the sound pressure magnitude spectrum of the object audio signal and a reference magnitude spectrum . For a single object audio signal, a frequency response is determined by:

]

**Scene-based audio**

The magnitude spectrum of the scene-based audio signal is determined by evaluating the root mean square of the magnitude response of [the Ambisonics coefficients , i.e.,

]  
Letters and respectively denote Ambisonics degree and index. The factor renormalizes the SN3D-normalized coefficients to N3D. The N3D-normalized root-mean-square of the coefficient-domain magnitude spectrum is equivalent of determining the root-mean-square in the spatial domain evaluated on a uniform sampling grid (Parseval’s theorem).

The UE frequency response is determined by dividing the magnitude spectrum by a reference measurement in order to compensate for the impact of the test signal and the measurement system. The frequency response is determined by

**Metadata-assisted spatial audio**

Frequency spectrum of the metadata-assisted spatial audio signal is defined as a ratio of the sound pressure magnitude spectrum of the transport channels (): and a reference magnitude spectrum . Letter denotes transport channel number. Thus, for each MASA transport channel a frequency response is determined by:

5.6.4 Directional information (single source)

5.6.4.1 Test method

The default arrangement for a single sound source is used, as defined in clause [**Error! Reference source not found.**] for each UE type. The decoded and rendered output format shall be the same as the IVAS audio format used by the UE. In addition to the default sound source direction, *L=[N]* directionsshall be evaluated, as indicated in Table X. In case the setup is realized with a turntable, the device shall be rotated around the vertical center of the UE.

**Table X: Additional source positions**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **fi** [deg] | **qi** [deg] | | -90° | 0° | | -45° | 0° | | 0° | 0° | | 45° | 0° | | 90° | 0° | | |  |  | | --- | --- | | **fi** [deg] | **qi** [deg] | | -90° | 0° | | -60° | 0° | | -30 | 0° | | 0 | 0° | | 30° | 0° | | 60° | 0° | | 90° | 0° | |

or

For each sound source position *(ii* *i*=1,...,L , the following procedure shall be used:

1. The test signal to be used for the measurements shall be [the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [xx], calibrated to an active speech level according to **Error! Reference source not found.**.]
2. The UE under test and the reference client are connected and configured as described in the clause [**Error! Reference source not found.**].
3. **Acoustical Interface:** The UE is mounted as described in clause [**Error! Reference source not found.**] and the acoustic source is positioned such that the source direction under test is met. The test signal is played via the acoustic source.

**Electrical Interface:** The test signal is generated by virtually placing the acoustic source such that the source direction under test is met as described in clause [**Error! Reference source not found.**].

1. The output format dependent directional metric calculations shall be done as defined in [4.4.4] for the tested format.

5.6.4.2 IVAS format specific definitions

**Stereo**

1. The left and right channel signals () are recorded by the test equipment.
2. The inter-channel time difference () is determined as described in Annex X between the signals and .
3. The inter-channel level difference (ICLD) is determined as follows:
   * 1. The active speech level according to Recommendation P.56 [xx] is calculated separately for and , resulting in and .
     2. ICLD for source position is calculated as:

1. The equivalent level difference for source position is calculated as:
2. The estimated source position in the stereo panorama is calculated as:

**Scene-based audio**

Direction of arrival analysis with scene-based audio shall be done as follows:

1. The B-format scene-based audio format representation is captured by the test equipment.
2. The intensity parameter is calculated from the B-format capture using the equation:

NOTE: The intensity is calculated in frequency domain and per subframe. Further steps are thus performed with subframe accuracy.

1. The direction of arrival estimation is calculated based on the intensity parameter using the equations:

,

,

Where the arctan function is assumed to be the computational variant “atan2” that solves the correct quadrant automatically

1. The estimated direction of arrival *(estest* is compared to the ground truth angle *(ii*.

[Editor’s note: Potentially in several frequency bands and potentially time averaged. Weighting could be done similarly as in MASA case by estimating subframe energies and energy ratios.]

If the sending UE is properly implemented in terms of directionality, phase and scaling of Ambisonics components, the DOA metric is expected to correspond to the ground truth angle. The DOA angle calculated from the Ambisonics components from the UE capture system shall be within some tolerances w.r.t. the ground truth angle to the incident sound.

**Metadata-assisted spatial audio**

Direction of arrival analysis with metadata-assisted spatial audio shall be done as follows:

1. The MASA format representation is captured by the test equipment. The MASA representation includes estimated source angles and direct-to-total energy ratio quantities per time-frequency tiles. Total energy per time-frequency tile is estimated from the transport signals.
2. The direct-to-total ratio times energy weighted azimuth and elevation angles (in radians) are mapped into Cartesian coordinate vectors , and over all subframes and [N] frequency bands:

where is the index of the frequency bands and is the index of the subframes.

1. The direction of arrival estimation *(estest* in degrees is calculated based on the mapped Cartesian coordinate vectors using the equations:

,

,

Where the arctan function is assumed to be the computational variant “atan2” that solves the correct quadrant automatically

1. The estimated direction of arrival is compared to the ground truth angle *(ii*.

[Editor’s note: Potentially in several frequency bands and potentially time averaged.]

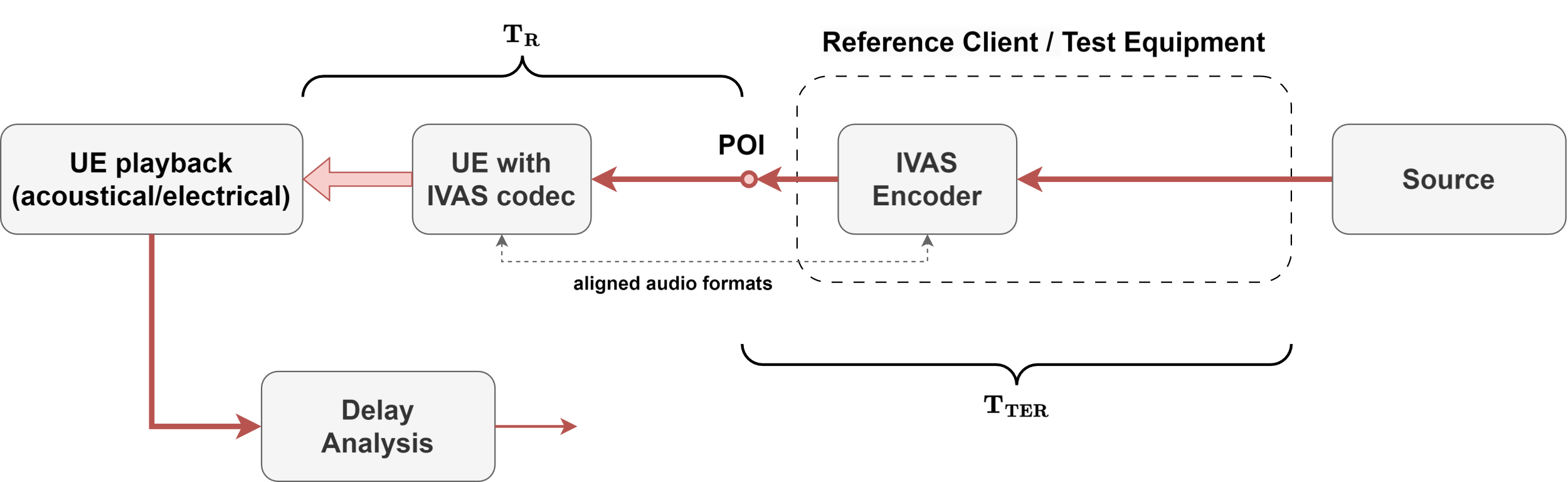
If the sending UE is properly implemented in terms of directionality and the energy ratio analysis for the MASA metadata, the DOA metric is expected to correspond to the ground truth angle. The DOA angle calculated from the MASA metadata from the UE capture system shall be within some tolerances w.r.t. the ground truth angle to the incident sound.

5.7 Test methods for receiving direction

5.7.1 Delay

**Test method**

The assessment of receiving UE delay is illustrated in Figure 1. The default arrangement (acoustical or electrical interface) in receiving direction as defined in clause **Error! Reference source not found.** for each UE type is used.

****

**Figure 1: Test setup for receiving delay**

1. The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [xx].
2. For the delay measurement in receiving, virtual positioning at azimuth 0° and elevation 0° according to clause [3.4.2] is used for the generated test signal.
3. The source signal is calibrated to a [level/loudness] of [-26 LKFS] as defined in sub-clause 5.5.1.
4. The UE is setup according to the clause [2.2]. The UE and the reference client connection is setup according to clause(s) [3.2], the source signal is encoded by the reference client, and inserted at the POI to the UE.
5. The capture of the UE output is carried out via …
   1. acoustical interface (headphones or loudspeakers): recording via diffuse-field equalized HATS.
   2. electrical interface: recording via corresponding reference interface. If the captured audio format is not stereo or binaural, the default IVAS binaural renderer is used to generate a binaural signal. To calibrate the digital signal into the (pseudo-)acoustical domain, it is assumed that a level of -26 dBov corresponds to an acoustical level of 73 dBSPL (-21 dBPa).

6) The UE delay in receiving direction is obtained between the electrical POI of the test equipment and the recorded signals at both ears. The cross-correlation analysis described in Annex X is carried out for both ear signals, using the mono source signal as reference. To obtain the overall delay TR, the results for left and right ears are averaged.

7) The measured delay shall be compensated by the delay TTER introduced by the test equipment. If applicable, also the binaural renderer for the electrical interface is compensated.

5.7.2 Loudness

5.7.2.1 Test method

1. The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [xx]
2. The source signal is calibrated to a [level/loudness] of [-26 LKFS] as defined in sub-clause 5.5.1.
3. The UE is setup according to the clause [2.2]. The UE and the reference client connection is setup according to clause(s) [3.2], the source signal is encoded by the reference client, and inserted at the POI to the UE.
4. The capture of the UE output is carried out via …
   1. acoustical interface (headphones or loudspeakers): recording via diffuse-field equalized HATS.
   2. electrical interface: recording via corresponding reference interface. If the captured audio format is not stereo or binaural, the default IVAS binaural renderer is used to generate a binaural signal. To calibrate the digital signal into the (pseudo-)acoustical domain, it is assumed that a level of -26 dBov corresponds to an acoustical level of 73 dBSPL (-21 dBPa).
5. The LLR in phon is calculated according to clause 8.3.3 of Recommendation ITU-T P.700 with the captured or rendered binaural signal.
6. The same binaural signal should be used to calculate Receiving Loudness Rating (RLR) according to Recommendation ITU-T P.79 [xx] for comparison to 3GPP TSs 26.131 [9]/26.132 [10].
   1. The inverse diffuse-field correction according to Recommendation ITU-T P.58 is applied on left and right channel of the recording to obtain the signal at DRP. Then DRP-to-ERP correction is applied.
   2. The reference signal used for the RLR calculation is the original test signal specified in step 1), calibrated to ‑16 dBm0.
   3. The RLR is calculated according to clause 8.2.3.2 of 3GPP TS 26.132.
7. Steps 2-7 should be repeated for additional source positions.

5.7.2.2 IVAS format specific definitions

**Stereo**

[

The test signal shall be the same test signal as defined in the 5.1.3 for both stereo channels.

]

**Object-based audio**

The virtual positioning of the source signal is done as defined in clause [3.4.2].

[Editor’s note: Test cases with rotated HATS TBD]

**Scene-based audio**

The virtual positioning of the source signal is done as defined in clause [3.4.2]. For the test cases with rotated HATS, the signal is rotated by multiplication with .

**Metadata-assisted spatial audio**

The virtual positioning of the source signal is done as defined in clause [3.4.2]. The test signal shall be the same signal as defined in clause 5.1.3 for each transport channel.

[Editor’s note: Test cases with rotated HATS TBD]

**Multichannel**

[

The test signal shall be… / The source position is set by… / TBD

]

5.7.3 Frequency response (single source)

[clause 5.2 of PDoc]

5.7.3.1 Test method

**Table X: Source positions for sensitivity/frequency characteristics**

|  |  |
| --- | --- |
| **Source azimuth** | **Source elevation** |
| 0 | 0 |
| 180 | 0 |
| 0 | 90 |
| 90 | 0 |
| -90 (270) | 0 |

The sensitivity/frequency characteristics may in addition be measured and reported for other positions.

[Editor’s note: Where it is possible to assign a certain distance to the object, a large value should be specified, to avoid corner cases with close distances]

The following procedure shall be used:

1. The test signal to be used for the measurements shall be as described in TS 26.132 clause [9.4.2 (SWB)].
2. The source signal is calibrated to a [level/loudness] of [-26 LKFS] as defined in sub-clause [3.4].
3. The UE and the reference client are setup according to clause(s) [3.2], the source signal is encoded by the reference client, and inserted at the POI to the UE.
4. The sensitivity/frequency characteristics are measured as described in TS 26.132 and are reported for the left and the right sides.

5.7.3.2 IVAS format specific definitions

**Stereo**

[

The test signal shall be the same test signal as defined in the 5.2.4 for both stereo channels.

]

**Object-based audio**

The virtual positioning of the source signal is done as defined in clause [3.4.2].

[Editor’s note: Test cases with rotated HATS TBD]

**Scene-based audio**

The source position of the test signal is set by presenting the encoder with a multi-component Ambisonics signal that represents a source from the particular incidence angle (see sub-clause 5.1.4).

**Metadata-assisted spatial audio**

The virtual positioning of the source signal is done as defined in clause [3.4.2]. The test signal shall be the same signal as defined in clause 5.2.4 for each transport channel.

[Editor’s note: Test cases with rotated HATS TBD]

**Multichannel**

[

The test signal shall be… / The source position is set by… / TBD

]

5.7.4 Inter-aural differences for binaural rendering

[clause 5.3 of PDoc. Suggest to use inter-aural (referring to binaural) instead of inter-channel (which would rather refer to stereo, multichannel, etc.?)]

[clause 5.4 of PDoc – could maybe be merged with ITD?]

5.7.4.1 Test method

If the UE supports head tracking, the test is performed for the reference direction with no HATS rotation and additionally for rotated HATS orientations. For the electrical interface test, the orientation information shall be passed to the electrical interface. For the acoustical test, HATS rotation shall be realized as described in clause [**Error! Reference source not found.**]. The source directions from Table X are rotated in the same way as the HATS is rotated, so that for all UE rotations the binaural signal generated by the UE shall only differ within the required tolerance.

**Table X: Source positions for inter-channel time difference measurement**

|  |  |
| --- | --- |
| **Source azimuth** | **Source elevation** |
| 0 | 0 |
| 180 | 0 |
| 0 | 90 |
| 90 | 0 |
| -90 (270) | 0 |

The following procedure shall be used:

1. The UE and the reference client are setup according to clause(s) [3.2], the source signal is encoded by the reference client, and inserted at the POI to the UE. The audio input format and bitrate shall be reported.
2. The left and right headphone/headset audio outputs from the UE are connected to the test system electrically, or acoustically using headphones and an ITU-T P.58 compliant head and torso simulator with associated left and right artificial ears as described in [2.2]. [Editor’s note: headtracking shall also be considered. Text TBD]
3. The volume control is set to nominal [Editor’s note: it is expected that the generic clauses of this specification will state that the volume control, unless otherwise stated, is set to meet the nominal RLR=8 +-3dB for each ear. The sentence “The volume control is set to nominal” may then be superfluous.].
4. The test signal is a CS-signal complying with ITU-T Recommendation P.501 using a PN-sequence with a length, T, of 4 096 points (for a 48 kHz sample rate test system). The duration of the complete test signal is as specified in ITU-T Recommendation P.501. [Editor’s note: this is the same signal as in TS 26.132 clause 7.5.4] The source signal is calibrated to a [level/loudness] of [-26 LKFS] as defined in sub-clause [3.4].

Editor’s note: The impact of codec on the test signal needs to be verified before performing the measurements.

1. If the UE supports head-tracking, the following horizontal plane HATS orientations φ0 are to be tested: φ0=0°,-30°, +30°.
2. For each simulated source position *(ii* *i*=1,...,L and each HATS orientation under test, the following procedure is repeated:
   1. The HATS is oriented according to the current orientation under test.
   2. The test signal is rotated just like the HATS orientation and is played to audio input of the refence client [the signal is proposed to be identical to TS 26.132 clause 8.5.4]. For each sub-test, the source position is set as defined in [5.4.5].
   3. The left and right headphone audio signals from the UE are captured electrically or acoustically, the capture method shall be reported. The analysis window shall include the PN-sequence part of the CSS signal. The correct positioning of the analysis window is accomplished by correcting for the delay of the test system and the particular UE, which is measured priorly.
   4. The left and right levels for the frequency bands of interest is noted.
   5. The transfer function between the left and the right channel is estimated [Editor’s note: details to be defined] and the inter-channel group delay is calculated from the phase response, as , where is the phase and is the angular frequency. The group delays for the different frequency bins are averaged from 200 to 2000Hz to obtain a single-figure inter-channel time difference.
3. The measurement is repeated for all source angles as listed in Table X.

5.7.4.2 IVAS format specific definitions

**Stereo**

Void.

**Object-based audio**

The virtual positioning of the source signal is done as defined in clause [3.4.2].

[Editor’s note: Test cases with rotated HATS TBD]

**Scene-based audio**

The virtual positioning of the source signal is done as defined in clause [3.4.2]. For the test cases with rotated HATS, the signal is rotated as defined in 5.1.4.

**Metadata-assisted spatial audio**

The virtual positioning of the source signal is done as defined in clause [3.4.2]. The test signal shall be the same signal as defined in 5.4.4 for each transport channel.

[Editor’s note: Test cases with rotated HATS TBD]

**Multichannel**

[

The test signal shall be… / The source position is set by… / TBD

]

Annex C (normative):  
Cross-correlation analysis

The following analysis method is used to determine the time difference (delay) between two time-discrete signals and by applying segmental cross-correlation with period T (in samples) and overlap L (in percent).

If not specified otherwise, a sampling rate of 48 kHz is assumed.

The envelope of the segmental cross-correlation function between and is calculated by means of the Hilbert transformation:

Each segment has a duration of T samples, using an overlap of L percent. The time difference is then determined by the time lag that provides the maximum value.

For expected shorter time differences (up to ~85 ms), T = 8192 and L = 50% are recommended.

For expected longer time differences (up to ~1.4 s), T = 131072 and L = 50% are recommended.