**3GPP TSG-SA WG4 Meeting #132S4-251020**

**Japan, Fukuoka, 19 – 23 May 2025 revision of S4-250307**

**Source: Rapporteur ATIAS\_Ph2[[1]](#footnote-1)**

**Title: Permanent Document ATIAS-2 v0.4**

**Agenda item: 7.7 / 14.2**

**Document for: Agreement**

# Introduction

The present document collects candidate changes to 3GPP TS 26.260 [2] on test methods for immersive UEs that were proposed during the work item ATIAS\_Ph2 [1]. If applicable, associated requirements for 3GPP TS 26.261 [3] are included as well.

The following clauses and subclauses are structured according to the objectives that are in scope of the ATIAS\_Ph2 WID:

* Consider additional requirements corresponding to the test methods in TS 26.260 and/or update requirements marked as TBD for sending and receiving characteristics of terminals in TS 26.261.
* Define new test methods and performance requirements/objectives, for the assessment of capture and playback of complex sound scenes, i.e., sound scenes with more than one source and from more than one defined direction.
* Define new test methods and performance requirements/objectives for the assessment of acoustic echo control. Test methods may be either completely new or be based on existing ones for mono telephony (from e.g., TS 26.132). In the latter case, it has to be investigated if and how such methods can be adapted for UEs providing immersive audio playback and/or capture capabilities.
* Define test methods and performance requirements/objectives for the assessment of binaural rendering in receive direction, including headtracking and motion-to-sound latency. Electrical as well as acoustical interfaces should be considered.
* Consideration of aspects, that are based on other ongoing work items, such as complexity level definitions in IVAS\_Codec\_Ph2.

NOTE: Updates to TS 26.259, in support of the development of the objective tests, are out of scope of this permanent document.

Each candidate change (i.e., updates to existing text or new clauses) is described in a separate subclause and contains:

- a brief summary and status of the proposed change

- a list of TDocs that initially proposed and further supported the change, as well as other related inputs

- all necessary changes to 3GPP TS 26.260 [2] and/or 3GPP TS 26.261 [3]. To reduce overhead and workload during drafting, the changes are prepared for a direct transfer into formal CRs.

This working procedure was preferred by the SA4 Audio SWG over continuously developed CRs, as it allows to add and track complementary information, avoiding also "changes over changes" in the history of the permanent document.

# References

The following references are used in the present document, but not in the candidate changes.

[1] [3GPP SP-241314](https://www.3gpp.org/ftp/tsg_sa/TSG_SA/TSGS_105_Melbourne_2024-09/Docs/SP-241314.zip), "WID on Terminal Audio quality performance and Test methods for Immersive Audio Services, Phase 2," HEAD acoustics GmbH, Qualcomm Incorporated, Nokia Corporation, Dolby Sweden AB, Orange, Xiaomi, Melbourne, 2024.

[2] [3GPP TS 26.260](https://www.3gpp.org/ftp/Specs/archive/26_series/26.260/26260-i10.zip): "Objective test methodologies for the evaluation of immersive audio systems", V18.1

[3] [3GPP TS 26.261](https://www.3gpp.org/ftp/Specs/archive/26_series/26.261/26261-i00.zip): "Terminal audio quality performance requirements for immersive audio services", V18.0

# Changes on references

The following references are not yet included in TS 26.260/261 but are used in the candidate changes in the present document.

[X1] Recommendation ITU-T G.122 (03/1993): "Influence of national systems on stability and talker echo in international connections".

[X2] ETSI ES 202 396‑1: "Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database".

[X3] IEC 61672-1: "Electroacoustics - Sound level meters - Part 1: Specifications".

[X4] Politis, A., & Gamper, H. (2017, October). Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays. In 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) (pp. 224-228).

# Changes on test setup

## UE types

### Summary

Specify test setups for UEs that are composed of multiple UE types (e.g. headset and handheld).

Progress: Initial discussion

### Related documents

|  |  |  |
| --- | --- | --- |
| TDoc | Title / Source | Notes |
| [S4-242015](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_130_Orlando/Docs/S4-242015.zip) | "On acoustic echo control testing and on UE type combinations", Nokia. | Identified the gap in UE type definitions |
| [SA4aA250003](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_Audio/Docs/SA4aA250003.zip) | "On testing simultaneous SND-UE-types", Nokia. | Further considerations regarding combined UE types |

### Changes for TS 26.260

Start change [

Editor's Note: If no further inputs are received, the approach of defining one of the involved UE types as the "primary one" could be a simple solution, which does not require new setup descriptions, figures, etc.?

] End change

# Changes on test methods and requirements

## Acoustic echo control

### Summary

Introducing test method for evaluating performance of acoustic echo control (AEC), which is adapted from existing specifications for IVAS-based UEs.

Progress: Initial proposal 🡪 stable test method, requirement is missing (SA4#132)

### Related documents

|  |  |  |
| --- | --- | --- |
| TDoc | Title / Source | Notes |
| [S4-242015](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_130_Orlando/Docs/S4-242015.zip) | "On acoustic echo control testing and on UE type combinations", Nokia. | Adapted test method for PDoc |
| [S4-241840](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_130_Orlando/Docs/S4-241840.zip) | "Test methods for UE acoustic echo control", HEAD acoustics GmbH. | Provided useful information on AEC test methods |
| [S4-250960](https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_132_Fukuoka/Docs/S4-250960.zip) | "On applied test signal for AEC assessment ", Nokia. | Improved test signal containing moving talkers |

### Changes for TS 26.260

Start change [

5.5.6 Acoustic Echo Control

One of the most relevant aspects for traditional as well as for immersive speech communication is that the generation and insertion of echo signals in send direction should be avoided wherever possible.

[The compressed British-English single talk sequence described in clause 7.3.3 of Recommendation ITU-T P.501 [14]] is used to simulate a conversation at the far-end. To evaluate echo cancellation performance with immersive audio playback/capture, the individual talkers are virtually arranged in different positions according to Table T1 and as visualized in Figure F1, including also dynamic movements of certain talkers.

Table T1: Source direction of each sentence in test signal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Talker | Time range [s] | | Source direction [°] | |
|  |  | From | To | From | To |
| 1 | M1 | 0.0 | 3.0 | 0 | 0 |
| 2 | M2 | 3.0 | 5.7 | 0 | 90 |
| 3 | M3 |  |  | 90 | 90 |
| 4 | F1 |  |  | 90 | -90 |
| 5 | F2 |  |  | -90 | -90 |
| 6 | F3 |  |  | 180 | 180 |
| 7 | F4 |  |  | 0 | 0 |
| 8 | F5 |  |  | 0 | 90 |
| 9 | F6 |  |  | 90 | 90 |
| 10 | M4 |  |  | 90 | -90 |
| 11 | M5 |  |  | -90 | -90 |
| 12 | M6 |  |  | 180 | 180 |

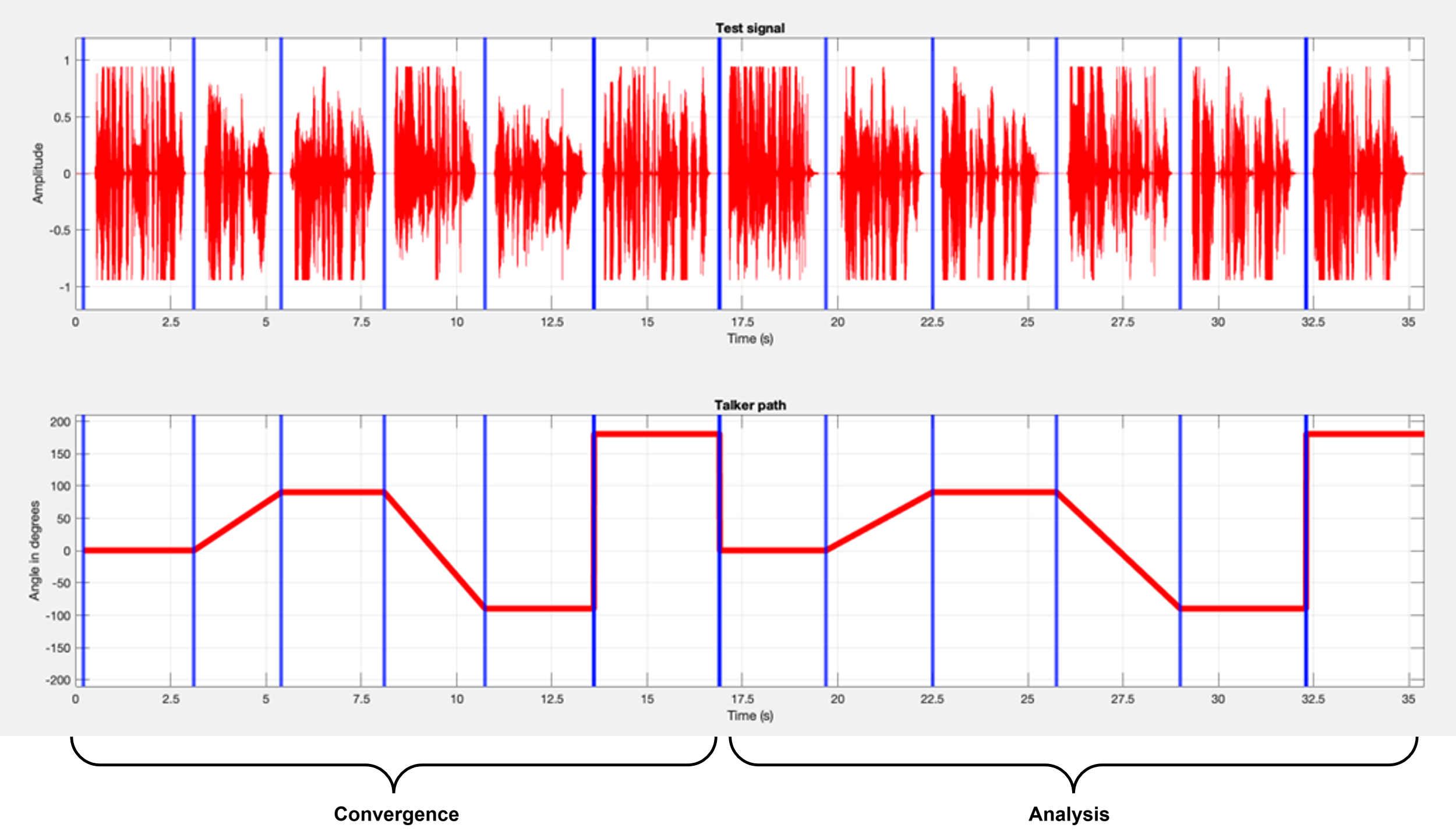


Figure F1: Sentence directions in the test signal

**Test method**

The calculation of terminal coupling loss (TCL) is based on the attenuation between the rendered input signal and rendered output signal versus frequency bands. For the rendering step, the IVAS renderer [30] shall be used. The following common measurement steps are applicable for all types of UE modes:

1) The source signal to be used for the measurements shall be [the compressed British-English single talk sequence described in clause 7.3.3 of Recommendation ITU-T P.501 [14].]

2) The test signal sentences shall be [virtually positioned according to clause 5.5.2. The source directions per sentence according to Table T1 shall be applied.]

3) The input signal shall be calibrated to [-20 LFKS]

4) The first 17.0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences).

5) The external renderer in the reference client shall be configured to [mono/stereo] output.

6) The analysis shall be conducted in 1/3-octave band intervals as given by the R.10 series of preferred numbers in ISO 3 [54] for each stereo audio channel. For the calculation, the averaged measured echo level at each frequency band is referred to the averaged, rendered test signal level measured in each frequency band, denoted as echo loss in the following.

7) TCL is calculated according to Annex B.4 of Recommendation ITU-T G.122 [X1], which utilizes tabulated data for any frequency range between and . a s unweighted echo loss from [100/300] Hz to [8/6.7] kHz by the following equations:

Where:

is the output/input power ratio at frequency Hz,

the ratio at frequency ,

the ratio at frequency  kHz.

Editor's Note / further considerations:

* Step 7) adapted from ITU-T P.381 (reference to " trapezoidal rule" is actually wrong), where also a different frequency range is specified.
* ambience or reverberation could be additionally simulated on the test signal

] End change

### Changes for TS 26.261

Start change [

Editor's note: Since the test method is adapted from mono telephony, the requirement on TCL could at least be drafted in the same way as e.g. in TS 26.131, just the single number (greater or less than 46 dB?) needs to be agreed.

] End change

## Headtracking & rendering

### Summary

Introducing test method for evaluating performance of headtracking and rendering capabilities of IVAS-based UEs.

Progress: Initial proposal / pending validation

### Related documents

|  |  |  |
| --- | --- | --- |
| TDoc | Title / Source | Notes |
| [SA4aA250001](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_Audio/Docs/SA4aA250001.zip) | "Test Method for Motion-to-Sound Latency", HEAD acoustics GmbH. | Provided proposal for M2S-latency measurement |

### Changes for TS 26.260

Start change [

4.0.2 Test equipment

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 [27].

For tests with head tracking, HATS rotation around the vertical axis should be realized using a HATS with motorized head rotation or comparable equipment consisting of P.57-compliant artificial ears mounted on a turntable or swivel apparatus (denoted as "motorized HATS" in the following clauses). For dynamic head-tracking measurements, the equipment is subject to angular velocity and noise emission requirements as specified in the respective test directives. Moreover, dynamic head-tracking measurements require the equipment to provide some kind of sensing mechanism for tracking orientation over time. For motorized or manual rotations of HATS and/or UE, error in orientation (elevation and azimuth) shall not exceed ±2°.

NOTE 1: A motorized rotation of HATS and/or UE is recommended even for static head-tracking measurements. Some UEs 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 after a span of time, e.g., to the current device orientation. This should be taken care of during the measurement. The measurement with the rotated HATS and/or UE should be performed quickly enough to prevent the reference direction from being spuriously readjusted. Therefore, it benefits from automation.

NOTE 2: Dynamic head-tracking measurements mimic the behavior of typical human head rotation. Angular rotation velocities and noise emission limits are required that may not be met by standard turntables.

[…]

5.7.5 Motion-to-Sound Latency

The test method applies to headset UEs providing headtracking functionality.

The test setup is as described in clause 5.4.3.2. A motorized HATS or comparable apparatus suitable for dynamic head-tracking measurements as described in clause 4.02 is used. The measurement apparatus shall be capable of realizing the below-stated movement trajectory. It shall feature a sensing mechanism that tracks the point in time at which the apparatus passes 0° orientation with a tolerance of [±2 ms]. Moreover, its noise emission shall not exceed [-70 dBPa] measured at the artificial ears after signal filtering. An example setup is illustrated in Figure X.

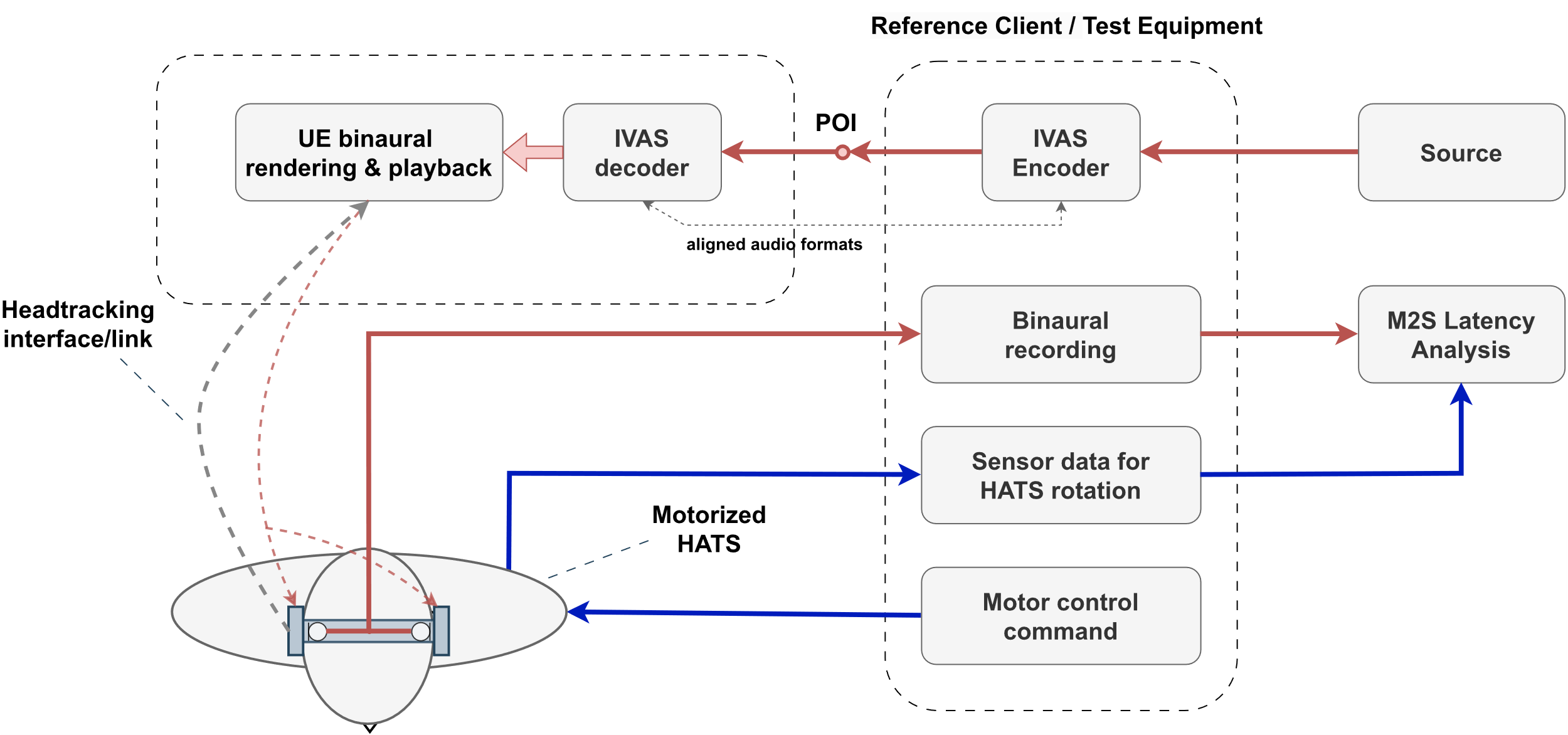


Figure X: Setup for measuring M2S latency using a motorized HATS

The test method evaluates interaural level differences that are calculated from the artificial ear signals with the following algorithm:

1) The left and right ear signals are each filtered with a narrow zero-phase bandpass filter that is designed to suppress noise of the measurement apparatus while not impacting the test signal. The filtered signals are denoted as and with sample index  and shall be made available at a sampling rate of .

2) The interaural level difference (ILD) is determined on the pre-processed left and right ear signals and , respectively. The ILD is calculated on short frames with index and can be stated as  
 .

For the calculation, frame length and hop size are used. The ILD is limited to  
, with .

3) The ILD is then smoothed over frames using a moving average with

Time instances per frame index are determined by

The motion-to-sound (M2S) latency measurement shall be conducted as follows:

1) The test signal to be used for the measurements shall be a sine signal at [2 kHz] virtually positioned to 0° azimuth and 0° elevation (frontal position for 0° head orientation) as described in clause 5.5.2. The signal shall be calibrated to [-20 LFKS] as described in clause 5.5.1.

2) The UE and the reference client are setup according to clauses 5.4.2 and 5.3.2, the source signal is encoded by the reference client and inserted at the POI to the UE. In case a receive volume control is provided, it shall be set to nominal setting as identified per clause 5.7.2.

3) The motorized HATS is oriented to 0°. For ILD calibration, the ear signals are captured for a duration of [3 s] without applying any motion to the UE. Then, is determined for the ILD calibration sequence and an average ILD (denoted as ) is calculated as the mean of .

4) The dynamic part of the measurement (in motion) is conducted as follows:

a) The motorized HATS rotates continuously and repeatedly between [+60°] (start orientation) and [-60°]. The 0° orientation shall be passed times, i.e., [6] times for each direction of rotation, so that a total of partial estimates of M2S latency are determined. The duration between two zero-crossings shall be less than [2.0 s].

b) In the angle range between [-10° and +10°], the angular velocity shall be constant with a nominal angular velocity of [108°/s].

c) During motion, the left and right ear signals and the time stamps (where the motorized HATS passes 0° orientation) are captured for .

5) After the last orientation is reached, the motorized HATS is immediately oriented to 0°. Step 3) is repeated to obtain the verification ILD, denoted as .

6) The M2S latency time is calculated as follows for each zero-crossing of the orientation:

a) The smoothed ILD versus time is determined.

b) The time stamp of the sound signal passing 0° orientation is determined by identifying the zero-crossing of .

c) The M2S latency for the -th measurement is determined by .

7) The overall M2S latency is determined by [calculating the median] of .

The principle of motion-to-sound analysis and determination of and is illustrated in Figure X+1.

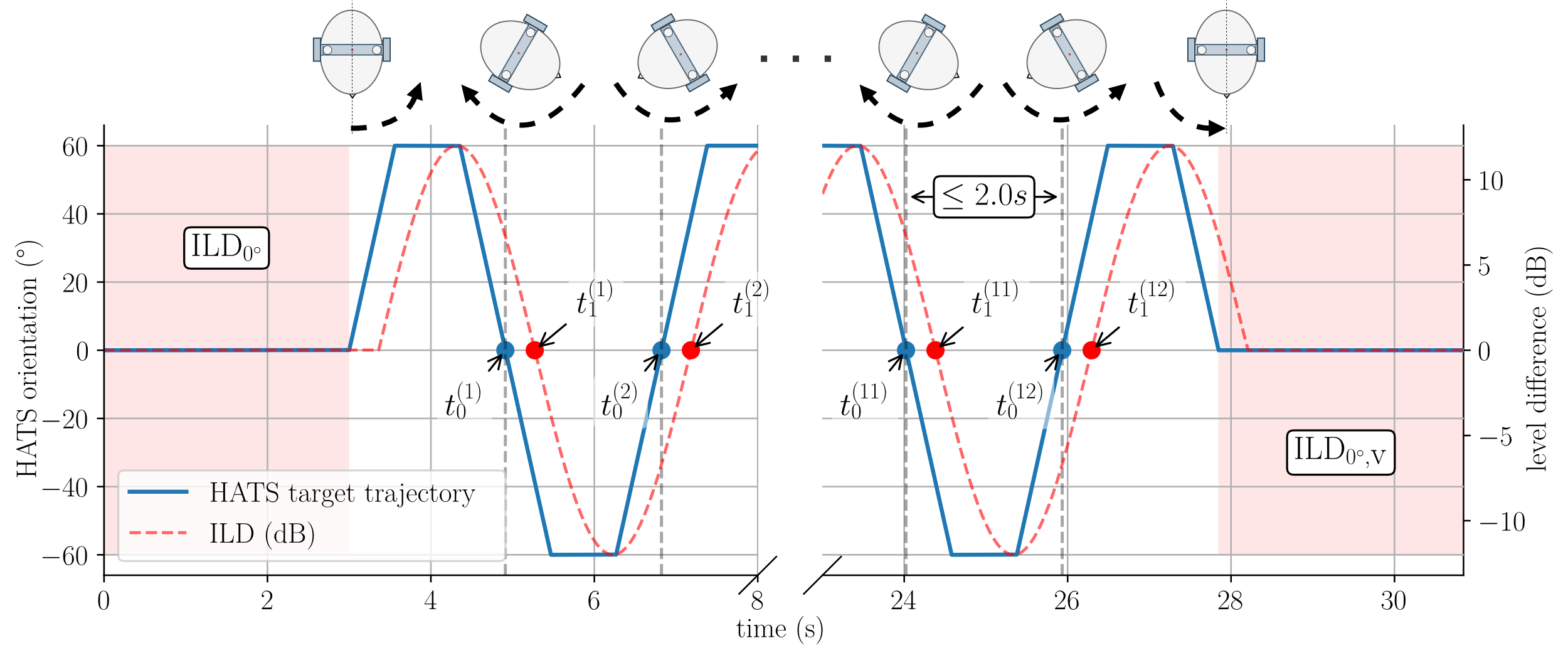


Figure X+1: Example setup for measuring M2S latency using a motorized HATS

Editor's notes/To Do:

- Outlier analysis of may help to detect erroneous measurements?

- The measurement method presented requires a constant reference/zero orientation. In some IVAS UE use-cases, a DUT may intentionally adapt its reference/zero orientation over time, for instance, based on the average of past orientations. The measurement method is not suited for such cases.

- The test method could also be extended to determine angular drift (with corresponding requirement for 26.261). This is important as a sufficiently low drift is needed to ensure a valid measurement of M2S latency.

- What about electrical interface testing (electrical interface intended for headset usage)?

] End change

### Changes for TS 26.261

Start change [

6.5.3 Motion-to-Sound Latency

The M2S latency measured with the test method from clause 5.7.5 from TS°26.260 should not exceed [80 ms] and shall not exceed [100 ms].

The absolute difference between the initial ILD calibration () and the verification ILD () shall not exceed [0.5 dB].

] End change

## Ambient sound field generation and transmission

### Summary

Following points have been identified to further assist on defining tests methods for assessing complex sound field transmission (e.g., ambience). Potential need for defining different capture types (in terms of noise suppression) in send directions:

* The expected performance of each capture type can be defined via sufficient set of test methods and respective requirements.
* Speech quality should be sufficient in every circumstance, including spatial characteristics.
* If the ambience is not targeted to be removed, the transmitted ambience should have similar characteristics as the original sound field.
* Flexible amount of ambience attenuation may be allowed, e.g., based on the user preference, noise conditions, use case, etc.
* If ambience is targeted to be removed, the amount of applied attenuation should be sufficient, while the quality and spatial characteristics of the active speech should be maintained
* Despite the targeted ambience attenuation, simultaneously received far-end signal should not cause interference with the captured and transmitted near-end signal.
* Possible presented capture types for further discussion:
  + Transparent capture (No noise suppression)
  + Nominal capture ("In between" No and Maximum noise suppressions)
  + Suppressed capture (Maximum noise suppression)
* Potential tests for assessing ambience transmission:
  + Loudness
  + Frequency response
  + Temporal characteristics
  + Spatial characteristics
* Common tests and requirements for assessing speech transmission are anticipated despite the applied capture type
* Capture type-specific tests and requirements may be needed to assess transmission of complex sound fields
* A need to robust way to identify applied capture type. Possible approaches:
  + "Transparency test": Based on the 5.6.2 of TS 26.260, SLR with noise(-like) signal to assess whether noise is transmitted or not. [S4aA250014]
  + SDP parameter to indicate and negotiate the applied capture type between UE and system simulator. [S4‑250597]
  + Manufacturer-defined capture type(s) supported by the UE
* In addition, a priority for identifying the applied capture types should be considered

Progress: Initial proposals.

### Related documents

|  |  |  |
| --- | --- | --- |
| **TDoc** | **Title / Source** | **Notes** |
| [S4-250205](https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_131_Geneva/Docs/S4-250205.zip) | "On testing UEs in ambience", Nokia | Initial views on UE ambience transmission testing, indicating need for different capture types |
| [S4aA250014](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_Audio/Docs/S4aA250014.zip) | "On transparency test", HEAD acoustics GmbH | A first concept for handling (at least two) anticipated UE behaviours in ambient noise conditions |
| [S4-250597](https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_131-bis-e/Docs/S4-250597.zip) | "On testing different capture types in send direction", Nokia | Further discussion on testing transmission of complex acoustic sound fields. Details on proposed SDP parameter presented. |
| [S4-250612](https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_131-bis-e/Docs/S4-250612.zip) | "Using a Periphonic Speaker Array for Improved Soundfield Reconstruction", Qualcomm, Incorporated | Equalization of periphonic loudspeaker array for symmetric microphone array according to ETSI TS 103 224. |
| [S4-251004](https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_132_Fukuoka/Docs/S4-251004.zip) | "Assessing Diffuseness Reproduction of Loudspeaker Layouts for ATIAS", Qualcomm, Incorporated | Equalization of periphonic loudspeaker array for spherical microphone array, comparing original and reproduced sound field with diffuseness metric. |

### Changes for TS 26.260

Start change [

3.1 Definitions

**Capture type**: Indication on the applied noise suppression type by the UE.

**Transparent capture type**: TBD

**Nominal capture type**: TBD

**Suppressed capture type**: TBD

Editor's note: Align naming convention ("capture type" and degrees) and definition with possible new SDP parameter for IVAS.

] End change

Start change [

5.5.3 Loudness calculation for immersive audio formats

The loudness level, K-weighted, relative to full scale (LKFS) for all immersive audio formats supported by the IVAS codec and captured at the POI in send direction is calculated according to ITU‑R BS.1770 [23]. The method supports stereo and all multichannel formats as input signals directly.

If the signal to be analysed is not already in one of these formats, the external IVAS reference renderer according to clause 6 of 3GPP TS 26.254 [30] shall be used to render the signal to 7.1+4 output format, which is then used for the loudness level calculation according to ITU‑R BS.1770 [23].] End change

Start change [

5.x.x Setup for ambient sound field simulation

Editor's Note: It is expected that a revision of ETSI TS 103 224 that is suitable for sufficiently accurate ambient sound field simulation will be made available for use in this specification.

[For headset UE, …

Table T2: Ambient sound fields used for headset UE

|  |  |  |
| --- | --- | --- |
| Ambient Sound Field | Duration | Averaged Sound Pressure Level of Microphone Array Elements |
| [Rock\_Concert.wav] | 30s | [95dBSPL(A)] |
| [Nature\_Scene.wav] | 30s | [40dBSPL(A)] |
| [Kindergarden.wav] | 30s | [60dBSPL(A)] |
| [Downtown\_Scene.wav] | 30s | [70dBSPL(A)] |

The HATS and the headset UE mounted to it are placed to the geometric centre of the loudspeaker array.

]

[For handheld hands-free UE, …

Table T3: Ambient sound fields used for hands-free UE

|  |  |  |
| --- | --- | --- |
| Ambient Sound Field | Duration | Averaged Sound Pressure Level of Microphone Array Elements |
| [Rock\_Concert.wav] | 30s | [95dBSPL(A)] |
| [Nature\_Scene.wav] | 30s | [40dBSPL(A)] |
| [Kindergarden.wav] | 30s | [60dBSPL(A)] |
| [Downtown\_Scene.wav] | 30s | [70dBSPL(A)] |

The handheld hands-free UE is placed in the geometric centre of the loudspeaker array.

]

[For table-mounted hands-free UE, …

Table T4: Ambient sound fields used for table-mounted UE

|  |  |  |
| --- | --- | --- |
| Ambient Sound Field | Duration | Averaged Sound Pressure Level of Microphone Array Elements |
| [Rock\_Concert.wav] | 30s | [95dBSPL(A)] |
| [Nature\_Scene.wav] | 30s | [40dBSPL(A)] |
| [Kindergarden.wav] | 30s | [60dBSPL(A)] |
| [Downtown\_Scene.wav] | 30s | [70dBSPL(A)] |

In accordance with clause 5.4.2.5, the table-mounted hands-free UE is placed in the geometric centre of the loudspeaker array.

]

For electrical interface UE with binaural input format, ambient sound field recordings according to Table T5 shall be used for electrical insertion, which can be found in clause 8.2 of ETSI ES 202 396‑1 [X2]. The binaural input signal shall be calibrated to [-26 LFKS], as described in clause 5.5.1.

Table T5: Ambient sound fields used for electrical interface UE and binaural input format

|  |  |  |
| --- | --- | --- |
| Ambient Sound Field | Duration | Averaged Sound Pressure Level of Microphone Array Elements |
| [Rock\_Concert.wav] | 30s | [95dBSPL(A)] |
| [Nature\_Scene.wav] | 30s | [40dBSPL(A)] |
| [Kindergarden.wav] | 30s | [60dBSPL(A)] |
| [Downtown\_Scene.wav] | 30s | [70dBSPL(A)] |

[For electrical interface UE with [other input] format(s), … TBD].

Editor's Note: Can we cover all UE types & coded formats here? Or are there "gaps" for which no ambient sound field simulation can be defined?

] End change

Start change [

5.4.4 Capture type configuration

UEs may support multiple capture types in terms of noise suppression in sending direction. All capture types supported by the UE shall be tested in send direction, which is selected according to the following priority:

1) Supported capture type(s) as indicated during negotiation in SDP

2) Supported capture type(s) as specified by the manufacturer

3) Capture type determined by the test method given in Annex TBD.

The tested capture type(s) shall be documented in the test report.

] End change

Start change [

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 [27]. 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 [28]. 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 and capture types in sending and receiving direction, as well as corresponding bitrate ranges.

- The reference client shall advertise via SDP the envisioned IVAS audio format and capture type 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, if not specified otherwise by any test method.

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

] End change

Start change [

5.6.5 Ambient Sound Field Loudness in Send Direction

[This test method only applies for UEs of capture type "transparent" (and/or "nominal"??)]

**Test method**

1) The UE under test and the reference client are connected and configured as described in clause 5.3.2. The UE is mounted as described in clause 5.4.2.

2) The ambient sound field generation as defined in clause 5.x.x is set up according to the corresponding UE type.

3) For acoustic measurements, the ambient sound field (test signal) is played back via the loudspeaker array.

For electrical interface, the test signal is played back via insertion into the electrical interface.

4) The captured signal is recorded at the POI and then decoded to the coded audio format of the UE.

5) The UE loudness level, K-weighted (LKFS) in sending direction is calculated according to clause 5.5.3.

6) The test is repeated for each ambient sound field in the corresponding tables of clause 5.x.x.

5.6.6 Ambient Sound Field Frequency Response in Send Direction

[This test method only applies for UEs of capture type "transparent" (and/or "nominal"??)]

**Test method**

1) The UE under test and the reference client are connected and configured as described in clause 5.3.2. The renderer in the reference client shall additionally be configured to mono output. The UE is mounted as described in clause 5.4.2.

2) The ambient sound field generation as defined in clause 5.x.x is set up according to the corresponding UE type.

3) For acoustic measurements, the ambient sound field (test signal) is played back via the loudspeaker array.

For electrical interface, the test signal is played back via insertion into the electrical interface.

4) The captured signal is recorded at the POI and the renderer in the reference client shall be configured to mono output.

5) The frequency response of the decoded and rendered output signal shall be calculated according to the format-specific definitions of the clause 5.6.3.2. For all formats, 1/12th octave intervals as given by the R40 series of preferred numbers in ISO 3 [6], for frequencies from 100 Hz to 12 kHz (inclusive) apply. The reference magnitude spectrum is the 1/12th octave spectrum of the mono downmix of the test signal.

6) The test is repeated for each ambient sound field in the corresponding tables of clause 5.x.x.

5.6.7 Ambient Sound Field Temporal Characteristics

[This test method only applies for UEs of capture type "transparent" (and/or "nominal"??)]

**Test method**

1) The UE under test and the reference client are connected and configured as described in clause 5.3.2. The renderer in the reference client shall additionally be configured to mono output. The UE is mounted as described in clause 5.4.2.

2) The ambient sound field generation as defined in clause 5.x.x is set up according to the corresponding UE type.

3) For acoustic measurements, the ambient sound field (test signal) is played back via the loudspeaker array.

For electrical interface, the test signal is played back via insertion into the electrical interface.

4) The captured signal is recorded at the POI and the renderer in the reference client shall be configured to mono output.

5) The decoded and rendered signal is referred to the mono downmix of the test signal as level versus time analysis according to IEC 61672 [X3], using an integration time of [250 ms]. The result represents the differences in temporal structure between the transmitted and reference ambient sound field.

6) The test is repeated for each ambient sound field in the corresponding tables of clause 5.x.x.

] End change

### Changes for TS 26.261

Start change [

5.3 Loudness

5.3.1 Loudness for single source

[…]

5.3.2 Ambient Sound Field Loudness in Send Direction

The Ambient Sound Field Loudness in Send Direction at the POI shall be according to the ranges in Table T6.

Table T6: Ambient sound fields used for electrical interface UE and binaural input format

|  |  |  |
| --- | --- | --- |
| Ambient Sound Field | Duration | Ambient Sound Field Loudness in Send Direction |
| [Rock\_Concert.wav] | 30s | Transparent Capture Type: [TBD to TBD] LKFS  Nominal Capture Type: [TBD to TBD] LKFS  Suppressed Capture Type: [TBD to TBD] LKFS |
| [Nature\_Scene.wav] | 30s | Transparent Capture Type: [TBD to TBD] LKFS  Nominal Capture Type: [TBD to TBD] LKFS  Suppressed Capture Type: [TBD to TBD] LKFS |
| [Kindergarden.wav] | 30s | Transparent Capture Type: [TBD to TBD] LKFS  Nominal Capture Type: [TBD to TBD] LKFS  Suppressed Capture Type: [TBD to TBD] LKFS |
| [Downtown\_Scene.wav] | 30s | Transparent Capture Type: [TBD to TBD] LKFS  Nominal Capture Type: [TBD to TBD] LKFS  Suppressed Capture Type: [TBD to TBD] LKFS |

Compliance shall be checked by the relevant tests described in TS 26.260.

5.4 Frequency response

5.4.1 Single Source Send Frequency response

[…]

5.4.2 Ambient Sound Field Frequency Response in Send Direction

5.4.2.1 Acoustical interface

The requirements for acoustical interface testing with **Transparent capture** are given in Table **T7**.

Table T7: Sending sensitivity/frequency requirements table for acoustical interface

|  |  |  |
| --- | --- | --- |
| Frequency (Hz) | Upper limit (dB) | Lower limit (dB) |
| 100 | TBD |  |
| 200 | TBD | TBD |
| 300 | TBD | TBD |
| 4000 | TBD | TBD |
| 8000 | TBD | TBD |
| 12000 | TBD |  |

The requirements for acoustical interface testing with **Nominal capture** are given in Table T8.

Table T8: Sending sensitivity/frequency requirements table for acoustical interface

|  |  |  |
| --- | --- | --- |
| Frequency (Hz) | Upper limit (dB) | Lower limit (dB) |
| 100 | TBD |  |
| 200 | TBD | TBD |
| 300 | TBD | TBD |
| 4000 | TBD | TBD |
| 8000 | TBD | TBD |
| 12000 | TBD |  |

No requirements for **Suppressed capture.**

5.4.2.2 Electrical interface

The requirements for acoustical interface testing with **Transparent capture** are given in .

Table T9: Sending sensitivity/frequency requirements table for acoustical interface

|  |  |  |
| --- | --- | --- |
| Frequency (Hz) | Upper limit (dB) | Lower limit (dB) |
| 100 | TBD |  |
| 200 | TBD | TBD |
| 300 | TBD | TBD |
| 4000 | TBD | TBD |
| 8000 | TBD | TBD |
| 12000 | TBD |  |

The requirements for acoustical interface testing with **Nominal capture** are given in .

Table T10: Sending sensitivity/frequency requirements table for acoustical interface

|  |  |  |
| --- | --- | --- |
| Frequency (Hz) | Upper limit (dB) | Lower limit (dB) |
| 100 | TBD |  |
| 200 | TBD | TBD |
| 300 | TBD | TBD |
| 4000 | TBD | TBD |
| 8000 | TBD | TBD |
| 12000 | TBD |  |

No requirements for **Suppressed capture.**

Compliance shall be checked by the relevant tests described in TS 26.260.

5.6 Temporal characteristics

**Transparent capture:** The differences [shall] be less than [± TBD dB]

**Nominal capture:** The range of differences [shall/should] be less than [± (TBD+3) dB]

**Suppressed capture:** The range of differences [shall/should] be higher than [± (TBD+3) dB]]

] End change

## Additional test methods for scene-based audio in sending direction

### Summary

Introducing additional test method for evaluating quality of scene-based audio in sending direction.

Progress: Initial proposal

### Related documents

|  |  |  |
| --- | --- | --- |
| TDoc | Title / Source | Notes |
| [S4-250282](https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_131_Geneva/Docs/S4-250282.zip) | "[ATIAS\_ph2] Scene-based audio capturing - Additional Test Methods", Bytedance. | Adapted test method for PDoc |

### Changes for TS 26.260

Editor's note: Some wording/terms/text fragments (highlighted in yellow) seems not very common/suitable for usage in a technical specification, also sometimes the usage of square brackets is inconsistent. This should be improved/rephrased before transferring it to the final CR.

Start change [

4.1.3 Diffuse Field Loudness

4.1.3.1 Test Setup

The recording environment is an anechoic chamber that meets clause 4.0.3. The test equipment should meet clause 4.0.2.

The following test material may be used:

[It’s still TBD on whether using the following material in separate test run or in a single test run.]

- [Uncorrelated pink noise, with length TBD. The loudness of each test signal is calibrated using the method [need clarification, need SPL measurement] in 3GPP TS 26.260.

- the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [14], which is calibrated as specified in clause 5.4.2 for the corresponding UE type and the tested capture mode (user or spatial).

- Music. Selection of material is TBD]

Sound source(s) are set to the following directions pointing to the device centre of the DUT to record candidate signals respectively. If the DUT is a headset UE, the DUT is mounted on a HATS. For tests with multiple sources, uncorrelated pink noise is used for each sound source.

The source directions are selected as follows.

1. Test 1-7: One single sound source is used, with its direction following all directions in Table TX1.

2. Test 8: Two sources at source directions No.2, and No.5 in Table TX1.

3. Test 9: Two sources at source directions No.3, and No.7 in Table TX1.

4. Test 10: Three sources at source directions No.1, No.4, and No.6 in in Table TX1.

Table TX1: Source directions

|  |  |  |
| --- | --- | --- |
| i | ϕi [°] | θi [°] |
| 1 | -90 | 0 |
| 2 | -60 | 0 |
| 3 | -30 | 0 |
| 4 | 0 | 0 |
| 5 | 30 | 0 |
| 6 | 60 | 0 |
| 7 | 90 | 0 |

For reference signals, we calculate the test material to Ambisonic, with the same order as recorded SBA, and the a-for-mentioned incident source directions:

where L is the number of active sound sources in the test scene, is the test material, and is the source direction.

Candidate signal is recorded by placing the center point of DUT at the center of the test environment, facing the No.4 sound source direction. The test signal is calibrated to a loudness of -26 dB LKFS. SBA time domain signals of the DUT is recorded as the candidate signal.

4.1.3.2 Test method

The W channel in the FOA recording is an omni-directional channel and hence is selected as the representative of the loudness of the diffuse field in a recording.

For each test, the normalized RMS amplitude of W channel is calculated, where is a short period of time. Then the loudness error is calculated by:

Example output: Loudness Error: 0.7%

4.1.4 Spatial Correlation

4.1.4.1 Test Setup

The recording environment is an anechoic chamber that meets clause 4.0.3. The test equipment should meet clause 4.0.2.

The following test material may be used:

[It’s still TBD on whether using the following material in separate test run or in a single test run.]

- [Uncorrelated pink noise, with length TBD. The loudness of each test signal is calibrated using the method [need clarification, need SPL measurement] in 3GPP TS 26.260.

- the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [14], which is calibrated as specified in clause 5.4.2 for the corresponding UE type and the tested capture mode (user or spatial).

- Music. Selection of material is TBD]

Sound source(s) are set to the following directions pointing to the device centre of the DUT to record candidate signals respectively. If the DUT is a headset UE, the DUT is mounted on a HATS. For tests with multiple sources, uncorrelated pink noise is used for each sound source.

The source directions are selected as follows.

1. Test 1-7: One single sound source is used, with its direction following all directions in Table TX2.

2. Test 8: Two sources at source directions No.2, and No.5 in Table TX2.

3. Test 9: Two sources at source directions No.3, and No.7 in Table TX2.

4. Test 10: Three sources at source directions No.1, No.4, and No.6 in in Table TX2.

Table TX2: Source directions

|  |  |  |
| --- | --- | --- |
| i | ϕi [°] | θi [°] |
| 1 | -90 | 0 |
| 2 | -60 | 0 |
| 3 | -30 | 0 |
| 4 | 0 | 0 |
| 5 | 30 | 0 |
| 6 | 60 | 0 |
| 7 | 90 | 0 |

For reference signals, we calculate the test material to Ambisonic, with the same order as recorded SBA, and the a-for-mentioned incident source directions:

where L is the number of active sound sources in the test scene, is the test material, and is the source direction.

Candidate signal is recorded by placing the center point of DUT at the centre of the test environment, facing the No.4 sound source direction. The test signal is calibrated to a loudness of -26 dB LKFS. SBA time domain signals of the DUT is recorded as the candidate signal.

4.1.4.2 Test Method

Spatial correlation, as defined as equation 28 in [X4], is used to evaluate the similarity between candidate signals and reference signals per Ambisonic order:

In which:

* m: SH degrees
* n: SH orders
* : Conjugate transpose (Hermitian) of the reconstructed SH coefficient vector for order n, degree m, and frequency f. This vector contains the reconstructed SH values across D measurement directions.
* : Diagonal weighting matrix of size D×D, containing sampling weights for the measurement grid directions. These weights approximate integration over the sphere to ensure orthonormality of the SH basis.
* : Ideal SH coefficient vector for order n and degree m, containing the theoretical SH values across the same D directions.
* : Energy of the reconstructed SH coefficients of order n, computed as
* : Energy of the ideal SH coefficients of order n, computed as

Example output:

Chart, line chart

Description automatically generated

Figure Y: Spatial Correlation of a recorded FOA vs. ground truth

4.1.5 Interaural Differences of Binaural Signals

4.1.5.1 Test Setup

The recording environment is an anechoic chamber that meets clause 4.0.3. The test equipment should meet clause 4.0.2.

The following test material may be used:

[It’s still TBD on whether using the following material in separate test run or in a single test run.]

- [Uncorrelated pink noise, with length TBD. The loudness of each test signal is calibrated using the method [need clarification, need SPL measurement] in 3GPP TS 26.260.

- the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [14], which is calibrated as specified in clause 5.4.2 for the corresponding UE type and the tested capture mode (user or spatial).

- Music. Selection of material is TBD]

Sound source(s) are set to the following directions pointing to the device centre of the DUT to record candidate signals respectively. If the DUT is a headset UE, the DUT should be worn by a HATS. For tests with multiple sources, uncorrelated pink noise is used for each sound source.

The source directions are selected as follows in Table TX3. Note that only one source is activated each time.

Table TX3: Source directions

|  |  |  |
| --- | --- | --- |
| i | ϕi [°] | θi [°] |
| 1 | -90 | 0 |
| 2 | -60 | 0 |
| 3 | -30 | 0 |
| 4 | 0 | 0 |
| 5 | 30 | 0 |
| 6 | 60 | 0 |
| 7 | 90 | 0 |

For reference signals, we calculate the test material to Ambisonic, with the same order as recorded SBA, and the a-for-mentioned incident source directions:

where L is the number of active sound sources in the test scene, is the test material, and is the source direction.

Candidate signal is recorded by placing the center point of DUT at the center of the test environment, facing the No.4 sound source direction. The test signal is calibrated to a loudness of -26 dB LKFS. SBA time domain signals of the DUT is recorded as the candidate signal.

4.1.5.2 Test Method

The recorded SBA is rendered into binaural cues for evaluating perceptual error objectively. The same render should be applied to the candidate signal and the reference signal, using the HRTF of HATS.

ITD and ILD of and are defined and calculated as follows.

- The ILD is defined as the difference between right and left levels (in dB) and is calculated for octave bands with center frequencies 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz. Octave bandpass filters according to IEC 61260-1 [25] shall be used. g)

- The ITD is defined as the difference in delay between right and left ear. For the calculation, the signal is prefiltered with a highpass filter with cut-off frequency 200 Hz and a lowpass filter with cut-off frequency 2000 Hz. Both filters shall use a filter order of four, which corresponds to a roll-off of 24 dB per octave. The ITD is calculated over the whole signal length using the cross-correlation method described in Annex C.

The ITD error is then calculated by:

and the ILD error of each octave band is calculated by:

Example result:

ITD error: -41

ILD error:

Table TX4: ILD Error Over Octave Bands

|  |  |
| --- | --- |
| Centre Frequency (Hz) | ILD error (dB) |
| 500 | 2.53 |
| 1000 | 9.07 |
| 2000 | 12.58 |
| 4000 | 7.42 |
| 8000 | 6.89 |

] End change

1. [Jan Reimes](mailto:jan.reimes@head-acoustics.com), HEAD acoustics GmbH [↑](#footnote-ref-1)