**3GPP TSG-SA WG4 Meeting #131S4-250082**

**Geneva, Switzerland, 17 – 21 February 2025**

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

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

**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".

# 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 |

### Changes

Start change [

] 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

### 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 |

### Changes

Change 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 different talkers are virtually positioned to alternating source directions according to Table T1 and as visualized in Figure F1.

Table T1: Source direction of each sentence in test signal

|  |  |  |
| --- | --- | --- |
| Direction | Talker | Source direction [°] |
| 1 | M1, M4 | 0 |
| 2 | M2, M5 | 45 |
| 3 | M3, M6 | -45 |
| 4 | F1, F4 | 90 |
| 5 | F2, F5 | -90 |
| 6 | F3, F6 | 180 |

A graph of sound waves

Description automatically generated

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 of each sentence shall be according to Table T1.]

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  Hz.

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.
* dynamic movement within the test signal sentences, or static directions per measurement
* ambience or reverberation could be additionally simulated on the test signal

] End change

## Headtracking & rendering

### Summary

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

Progress: Initial proposal

### 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

Change 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.

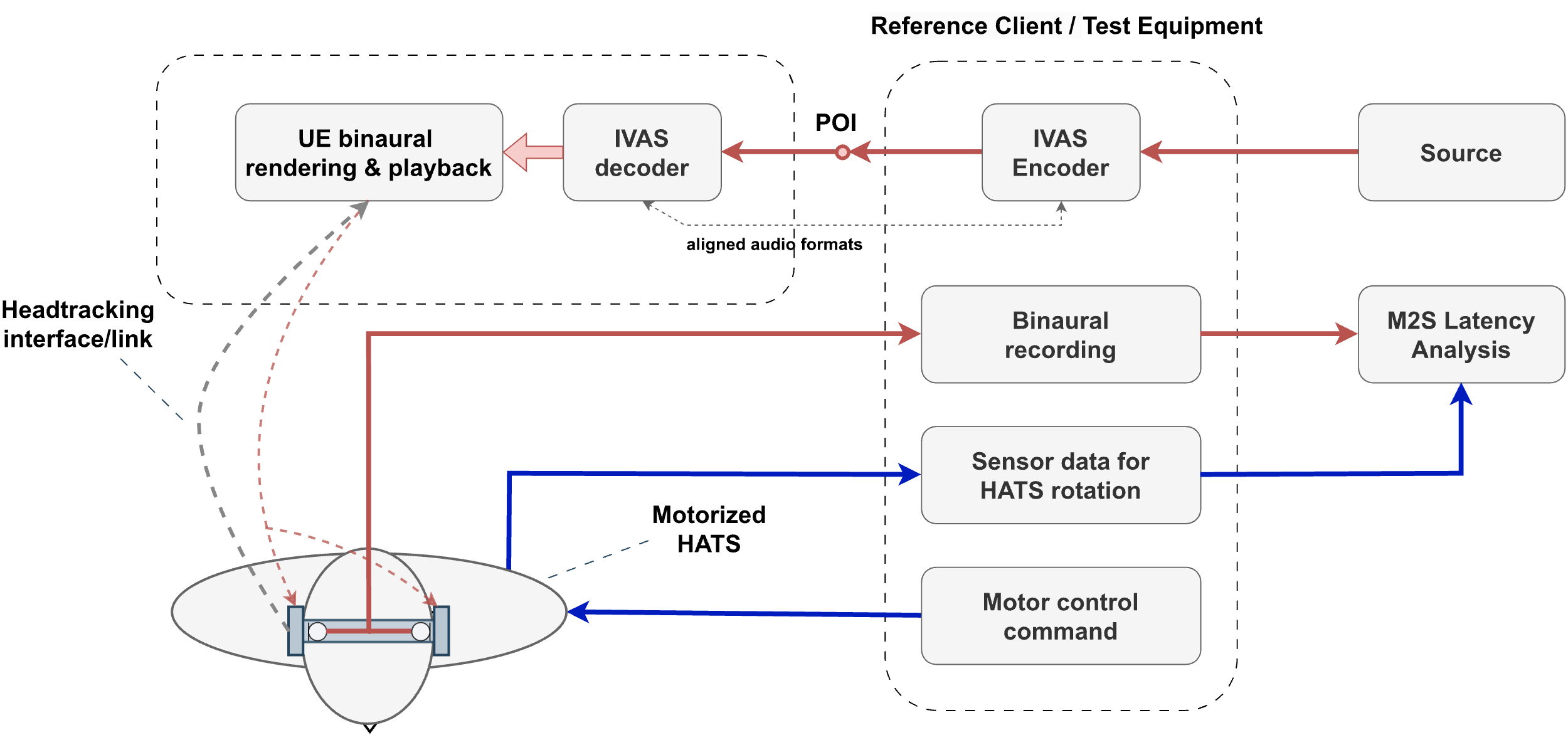


Figure X: Setup for measuring M2S latency using a motorized HATS

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.

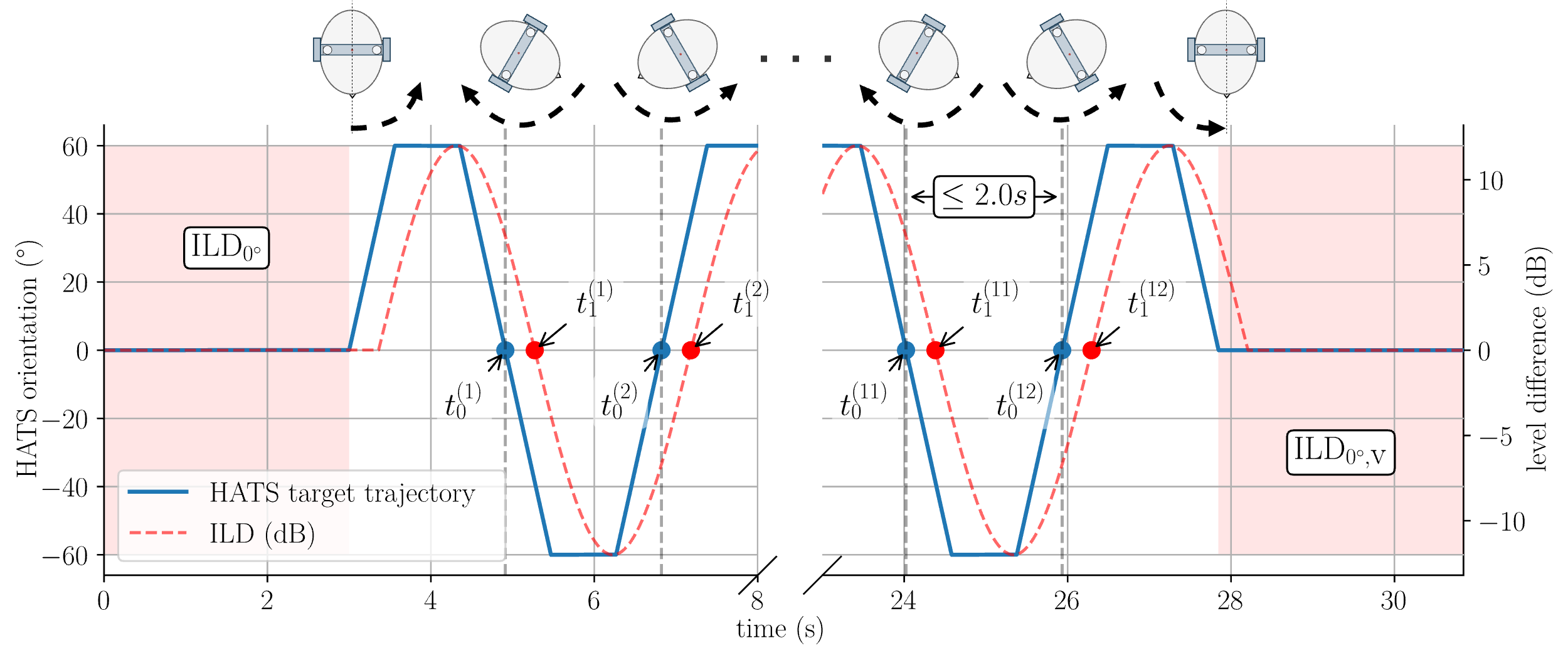


Figure X+1: Example setup for measuring M2S latency using a motorized HATS

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

Change 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

## Complex acoustic sound field 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
    - Nominal capture
    - Suppressed capture
* 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. Presented 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. [S4A250014]
  + SDP parameter to indicate and negotiate the applied capture type between UE and system simulator. [S4-250597]
* Following changes are to facilitate further changes in the specification, according to the above points.

Progress: Initial skeleton for required specification changes

### Related documents

|  |  |  |
| --- | --- | --- |
| **TDoc** | **Title / Source** | **Notes** |
| S4-250205 | "On testing UEs in ambience", Nokia. | Initial views on UE ambience transmission testing, required |
| S4A250014 | “On transparency test”, HEAD Acoustics Gmbh | A first concept for handling (at least two) anticipated UE behaviours in ambient noise conditions |
| S4-250597 | “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. |

### Changes

Change for TS 26.260:

**Start change [**

5.6.1 Single source tests

Editor’s note:

* All existing tests in send direction from section 5.6 are moved to here
* Test methods targeted to relevant capture characteristics for speech transmission, typically involving only speech as a test signal. Applicable for all send capture types

…

5.6.2 Complex sound field tests

5.6.2.1 Loudness (Ambience)

[TBD]

Editor’s note:

* Requires background noise simulation system to produce ambient sound field

5.6.2.2 Frequency response (Ambience)

[TBD]

Editor’s note:

* Possible changes expected to section 5.3 to address necessary changes for enabling different capture types
* Possible changes expected to sections 4.0. and 5.4. to address ambient sound field simulations method
* Possible changes expected to section 5.5 to address applied ambient sound field signals

**] End change**

Change for TS 26.261:

Start change [

…

5.3 Loudness

5.3.1 Loudness for single source

Editor’s note:

* Existing requirement from 5.3 are incorporated here
* Common requirements for all send capture type

…

5.3.2 Loudness for ambient source

[TBD]

Editor’s note:

* Capture type specific requirements expected according to the anticipated noise suppression behavior
* Certain loudness required, when ambience is expected to be transmitted
* Certain upper loudness limit expected, when ambience is expected to be suppressed
* Possible “in-between” behavior expected, where transmitted ambience is slightly attenuated, but still clearly audible

…

] End change

Start change [

5.4.2 Frequency response for ambient source

[TBD]

Editor’s note:

* May have capture type specific requirements

…

] End change

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