3GPP TSG-SA4#124 meeting Tdoc S4-231065

22-26 May 2023, Berlin, Germany Revision of S4-230645

**Source: Editor[[1]](#footnote-1) / Audio SWG**

**Title: Pdoc ATIAS-1 v0.4.0**

**Agenda Item: 14.1**

Table of Contents

[1 Introduction 2](#_Toc135924328)

[2 Test Configurations 2](#_Toc135924329)

[3 Test Conditions 2](#_Toc135924330)

[3.1 UE configuration 2](#_Toc135924331)

[3.2 Test Equipment 2](#_Toc135924332)

[4 Candidate sending side test methods and requirements 2](#_Toc135924333)

[4.1 Sending frequency response of captured Ambisonics components 2](#_Toc135924334)

[4.2 Sending directional response of captured Ambisonics componentss 3](#_Toc135924335)

[4.2.1 Angular-dependent sensitivity 3](#_Toc135924336)

[4.3 Direction of arrival estimation under free-field propagation conditions 4](#_Toc135924337)

[4.4 Directivity test of FOA using virtual microphones 6](#_Toc135924338)

[4.5 Scene-based audio spatial separation with two simultaneous acoustic sources in free-field propagation conditions and FOA decoding 9](#_Toc135924339)

[4.5.1 Definition 9](#_Toc135924340)

[4.5.2 Test conditions 9](#_Toc135924341)

[4.5.3 Measurement 10](#_Toc135924342)

[4.5.4 Annex A: Test signal and analysis for spatial separation with simultaneous acoustic sources 12](#_Toc135924343)

[4.6 Spatial separation for multiple acoustic sources based on multichannel output 16](#_Toc135924344)

[4.6.1 Test conditions 16](#_Toc135924345)

[4.6.2 Measurement 17](#_Toc135924346)

[4.7 Spatial perception test for stereo UE in ATIAS 19](#_Toc135924347)

[4.7.1 Definition 19](#_Toc135924348)

[4.7.2 Test Conditions 19](#_Toc135924352)

[4.7.3 Test Configurations 19](#_Toc135924353)

[4.8 Sending side audio performance assessment for Immersive Audio Systems in wind noise 22](#_Toc135924354)

[4.8.1 Introduction 22](#_Toc135924355)

[4.8.2 Recommendations for wind noise simulations for terminal testing 25](#_Toc135924356)

[5 Candidate receiving side test methods and requirements 26](#_Toc135924357)

[References 26](#_Toc135924358)

[Revision history 27](#_Toc135924359)

# Introduction

The ATIAS work item intends to specify test methods for objective characterization of terminals for 3GPP immersive services along with requirements. This Permanent Document collects candidate test methods and associated requirements that will form a pool out of which selected methods and requirements will be incorporated into TS 26.260 (Objective test methodologies for the evaluation of immersive audio systems) and, respectively, TS 26.261 (Terminal audio quality performance requirements for immersive audio services).

Several of the clauses below have retained explanations and examples from the original input documents. This may in some cases not be suitable for a specification text and it is expected that further editing will be necessary when incorporating the texts into the specifications.

# Test Configurations

[Editor’s note: Collect general test setups in this clause, similar to TS 26.132 clause 5. The individual test descriptions could then point to this clause.]

# Test Conditions

[Editor’s note: Collect general test conditions in this clause, similar to TS 26.132 clause 6]

[

## UE configuration

For testing the UE shall be configured for the relevant and/or envisioned use cases. During the tests potential internal development modes of the UE shall be disabled.

Ideally UEs should be test as is even if they have signal enhancement features. However, if performance issues are encountered and the UE allows to disable certain of these features like noise suppression, the tests should be repeated with these features disabled to document the possible cause of the problem.

[Editor’s note: Define IVAS codec operation mode(s) and bit rates to be used. Preferably such modes and bit rates should be used that have minimum impact on the acoustic tests.]

## Test Equipment

[Editor’s note: Requirements on acoustic test equipment, system simulator, reference client, …]

]

# Candidate sending side test methods and requirements

[

The following methods have been incorporated from [1]:

## Sending frequency response of captured Ambisonics components

*Definition:*

Ratio of the sound pressure magnitude spectrum of the DUT for Ambisonics component (, ): and a reference diffuse field sound pressure spectrum ( ). Letters and respectively denote Ambisonics degree and index. This means, for each Ambisonics component a sending frequency response is measured:

.

*Ideal characteristic:*

The ideal send frequency response of the captured Ambisonics components would be flat, i.e. be frequency independent (for plane wave input).

*How to formulate requirements:*

One conceivable way to measure against requirements can be taken from TS 26.260. A decorrelated pink noise test stimulus is simultaneously played over all speakers of a periphonic array. The reference diffuse field sound pressure spectrum is obtained through recordings with a diffuse-field microphone positioned at the geometric centre of the periphonic array. The sound pressure magnitude spectrum of the DUT for Ambisonics component (, ) is obtained through measuring with the DUT at the geometric centre and extracting and analysing that (B-format) component after coding, transmission and decoding as in 26.260 Figure 1.

The sensitivity/frequency response shall meet the ideal characteristics, within some tolerance.

## Sending directional response of captured Ambisonics componentss

### Angular-dependent sensitivity

*Definition:*

Microphone angular-dependent sensitivity can be described by (in Volt/Pascal or digital amplitude per Pascal), e.g., referring to the output voltage or digital level generated by a microphone for a given sound pressure at the microphone location for an incident plane wave from a certain direction . Likewise, the angular-dependent sensitivity of the capture of each Ambisonics component (, ) can be described as . As this sensitivity may also be frequency dependent, frequency is a further parameter:

.

Angular-dependent sensitivity of the capture of an Ambisonics component (, ) can also be defined for multiple () incident plane waves from respective directions . This yields:

with and .

Of interest may be the cases , and . For , the two directions should be 90 degrees apart (see [1-1]). For , a reasonable setup could be where the directions would correspond to the Fliege positions of the considered Ambisonics order .

*Ideal characteristic:*

The ideal angular-dependent sensitivity for the captured Ambisonics component (, ) is obtained from the spherical harmonics equations. For simplicity, it is assumed that the spherical coordinate systems of the measurement room and the captured Ambisonics signals are aligned.

For the case , we get

with being the real valued spherical harmonics of the degree and index with .

Thus, if the reference client offers output of this Ambisonics component, it will in the ideal case and valid single plane-wave assumption (e.g., a distant-enough point source to approximate plane waves at the UE) follow the above relation. The ideal frequency response characteristic is flat. Care needs to be taken to apply the proper Ambisonics component ordering and normalization.

For any , we get

Cases for K > 2 are not covered further.

*How to formulate requirements:*

The measurement for the simplest case of can be done according to the principles outlined above (under 2.1) using pink noise test stimuli. Unlike 2.1, only a single speaker at a time is used.

The angular-dependent sensitivity shall meet the ideal characteristics, within some tolerance.

**[**

## Direction of arrival estimation under free-field propagation conditions

**2.3.1 Definition**

Direction of arrival (DOA) is defined as the spherical angle pointing towards the sound source. DOA is relative to the capture device position. This measurement compares the DOA estimation with the ground-truth DOA under free-field propagation conditions.

Note: The DOA estimation performance in reverberant environments may be different and is not covered by this test.

**2.3.2 Test conditions**

**Free-field propagation conditions**

- The test environment shall contain a free-field volume, wherein free-field sound propagation conditions shall be observed.

- The free-field sound propagation conditions shall be observed down to a frequency of 200Hz.

**[Test environment noise floor]**

[Editor’s note: The test environment noise floor may not have to specified in this clause. Likely, a general clause for the whole specification is sufficient.]

**Loudspeaker array**

A real or simulated loudspeaker array comprising L loudspeakers located at a set of predefined directions *(ii* *i*=1,...,L , from the geometric center of the *loudspeaker array* shall be used.

**2.3.3.1 Measurement for Scene-based audio**

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

1. The UE under test is connected to a test system composed of a 3GPP wireless system simulator and reference client with an IVAS session established with B-format output. The codec shall be operated with scene-based input format at [512] kbit/s. The audio input format and bitrate shall be reported. The decoder/renderer option shall be FOA.
2. [TBD] test signal of [TBD s] length is played over the loudspeaker.

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

c) The B-format scene-based audio format representation is captured.

d) 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.

e) 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

f) 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.

**2.3.3.2 Measurement for Metadata-assisted spatial audio**

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

1. The UE under test is connected to a test system composed of a 3GPP wireless system simulator and reference client with an IVAS session established with metadata-assisted spatial audio format output. The codec shall be operated with Metadata-assisted spatial audio input format at [512] kbit/s. The audio input format and bitrate shall be reported. The decoder/renderer option shall be MASA.
2. [TBD] test signal of [TBD s] length is played over the loudspeaker.

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

c) The Metadata-assisted spatial audio format representation is captured. The MASA representation includes estimated source angles and energy-related quantities per time frequency tiles, which are further analysed as follows.

d) 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 frequency bands:

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

[Editor’s note: Signal length [TBD] = , where total number of subframes *K* = 1,2,…*k*,]

e) The direction of arrival estimation *(estest* 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

f) 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.

**2.3.4 Comments for DOA test method design**

Based on experimental evidence, the following points should be taken into an account:

- The DOA analysis based on analysing FOA and MASA signals were found to produce nearly equivalent results.

- Extreme angles should be handled accordingly. At the extreme angles, the absolute angle error can be very large, while the distance between the estimated sound source locations is small.

- (Mean) absolute angle error metric may not be the most applicable error calculation method. It should be examined whether e.g., spherical distance would be more suitable error metric

- Placement of the DUT is very critical for the test. It is suggested that the test should be performed multiple times, with replacement of the DUT between the measurements.

- Cone-of-confusion errors may occur in measurements of small devices with a limited number of microphones. These errors could be handled by e.g., limiting the measurement points or ignoring such errors if the number of errors is within acceptable limits.

- The applicability of the DOA test method for HOA capture is ffs.

**]**

[

## Directivity test of FOA using virtual microphones

*Definition:*

A virtual microphone is created by a linear combination of the ambisonics components. As an example, for FOA, a cardioid, super-cardioid, figure-of-eight etc microphone can be constructed and pointed in an arbitrary direction, by adding or subtracting portions of the four FOA ambisonics components.

It is called a *virtual* microphone because its characteristics are manipulated without affecting the microphone itself. Compare with music production; the mix engineer can *in post-production* select where to point a microphone, and even construct an infinite number of virtual microphones pointing in various directions with various beam widths. A simplified variant of FOA capture is MS stereo recording using one omni and one figure-of-eight capsule, where the stereo width is decided in post-production by turning (usually two) virtual microphones’ mutual angle and setting their pickup pattern to taste. For ATIAS, this means the test system can steer an infinite number of virtual microphones anywhere and controlling the width of the pickup beam, without interacting with the UE that is “just” supplying ambisonics components to the test system.

*Ideal characteristics*

An ideal first-order directional microphone’s directional pickup in a plane can be described as:

, where *p* =1 provides omni-directional, *p*=0.5 cardioid, *p* =0.37 super-cardioid, *p* =0.25 hyper-cardioid, *p* =0 bi-directional, etc.

This polar pattern can further be rotated, as illustrated below. The GIF illustrates adding portions of FOA components W, X and Y (“Pattern” corresponds to *p* above):

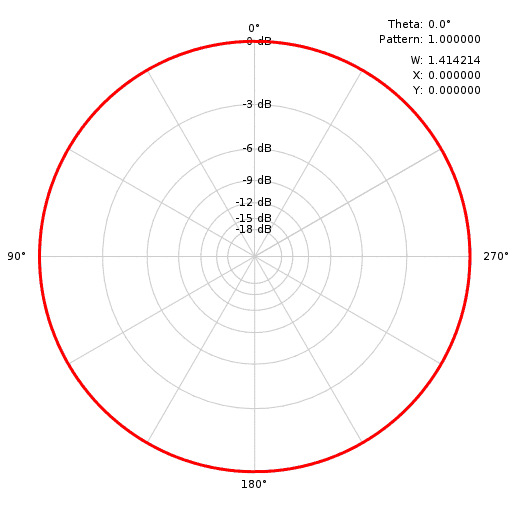


Figure 1 Illustration of beam and pattern steering (copied from [1-3])

With these simple means we are able to e.g. construct two simultaneous virtual microphones, one pointing at a sound source, the other pointing away from the sound source. This can be as simple as summing/subtracting the outputs from the reference decoder:

* Virtual mic A = W+Y
* Virtual mic B = W-Y

In the ideal case, if using two virtual back-to-back cardioids, one pointing towards a single sound source and one pointing away from it, only the first microphone shall produce an output signal.

*How to formulate requirements:*

Example: When a single sound source is placed at the positive Y direction in relation to the UE, the level of signal (W+Y) shall exceed the level of signal (W-Y) by at least X dB[[2]](#footnote-2).

This test can be expanded to cover a variety of incidence angles. Or it can be expanded to rotate the virtual microphones while keep a single source constant.

The merit of using two virtual microphones (which are preferably implemented simultaneously), instead of only one, is that the testing is robust to dynamic range processing in the UE, thus spatial aspects can be robustly assessed with this method. This is achieved by the virtual microphones “pointing” at two or several incidence angles while the UE is subjected to only one source from one angle. A merit compared to [1-1], is that we avoid two simultaneous sound sources and two separable signals.

The method is equally usable for HOA, by defining appropriate linear combinations for the ambisonics components from the decoder.

]

[

Editor’s note: the methods on sending spatial separation with two simultaneous sources and with multi-channel setup share the same basic principles and should be harmonized.

The following methods have been incorporated from [4]:

## Scene-based audio spatial separation with two simultaneous acoustic sources in free-field propagation conditions and FOA decoding

### Definition

The Scene-based audio sending spatial separation with two simultaneous acoustic sources measures the difference between the UE captured polar patterns of the Ambisonics B-format components and the expected ground-truth when the UE is subjected to two simultaneous acoustic sources at predefined directions, *(ii* *i*=1,...,L.

### Test conditions

**Free-field propagation conditions**

- The test environment shall contain a free-field volume, wherein free-field sound propagation conditions shall be observed.

- The free-field sound propagation conditions shall be observed down to a frequency of 200Hz.

**[Test environment noise floor]**

[Editor’s note: The test environment noise floor may not have to specified in this clause. Likely, a general clause for the whole specification is sufficient.]

**Loudspeaker array**

An array of coaxial loudspeakers is located at a set of predefined directions *(ii*, *i=1,…,6*, from the geometric center of the UE. The different locations may be realized using multiple loudspeakers or by rotation of at least two loudspeakers or by rotation of the UE.

The N=4 periphonic array can be augmented with these additional 5 positions (Y2 is already present).

In case the UE has motion compensation (automatic rotation of the captured soundfield depending on pose), physical rotation of the UE shall not be used to achieve the predefined directions and it shall be ensured that there is no misalignment of X, Y and Z axes due to the motion compensation.

[QCOM – Andre] Here again, the alignment with the periphonic array would be of help since the test lab would not have to worry about the issue above and no movement of the device would be required. See comments above.

[Editor’s note: this applies for general audio case. For communication HATS playback might be considered.]

The distance from the loudspeaker front baffle to the center of the UE shall be at least 1m and equal within ±[x]% for all loudspeakers. [Editor’s note: check if 1m is sufficient, considering the proximity effect at 200/275 Hz, which we would like to avoid as it biases the test result. This should go (later) into a separate clause]

The loudspeakers shall be equalized to a flat frequency response and equal sensitivity with the UE absent, using a [measurement microphone and diffuse-field equalization] placed at the UE position. The microphone shall point in the positive Z direction with its membrane in the XY plane.

Table X: Location of loudspeakers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Position | i | i [deg] | i [deg] | Comment |
| X1 | 1 | 0 | 0 | 0 deg (frontal) incidence to the DUT, along the X-axis |
| X2 | 2 | 0 | 180 | opposite to X1, along the X-axis |
| Y1 | 3 | 0 | -90 | -90 deg incidence to the DUT, along the Y-axis |
| Y2 | 4 | 0 | 90 | opposite to Y1, along the Y-axis |
| Z1 | 5 | 90 | 0 | “from the ceiling” incidence to the DUT, along the Z-axis |
| Z2 | 6 | -90 | 0 | opposite to Z1, along the Z-axis |

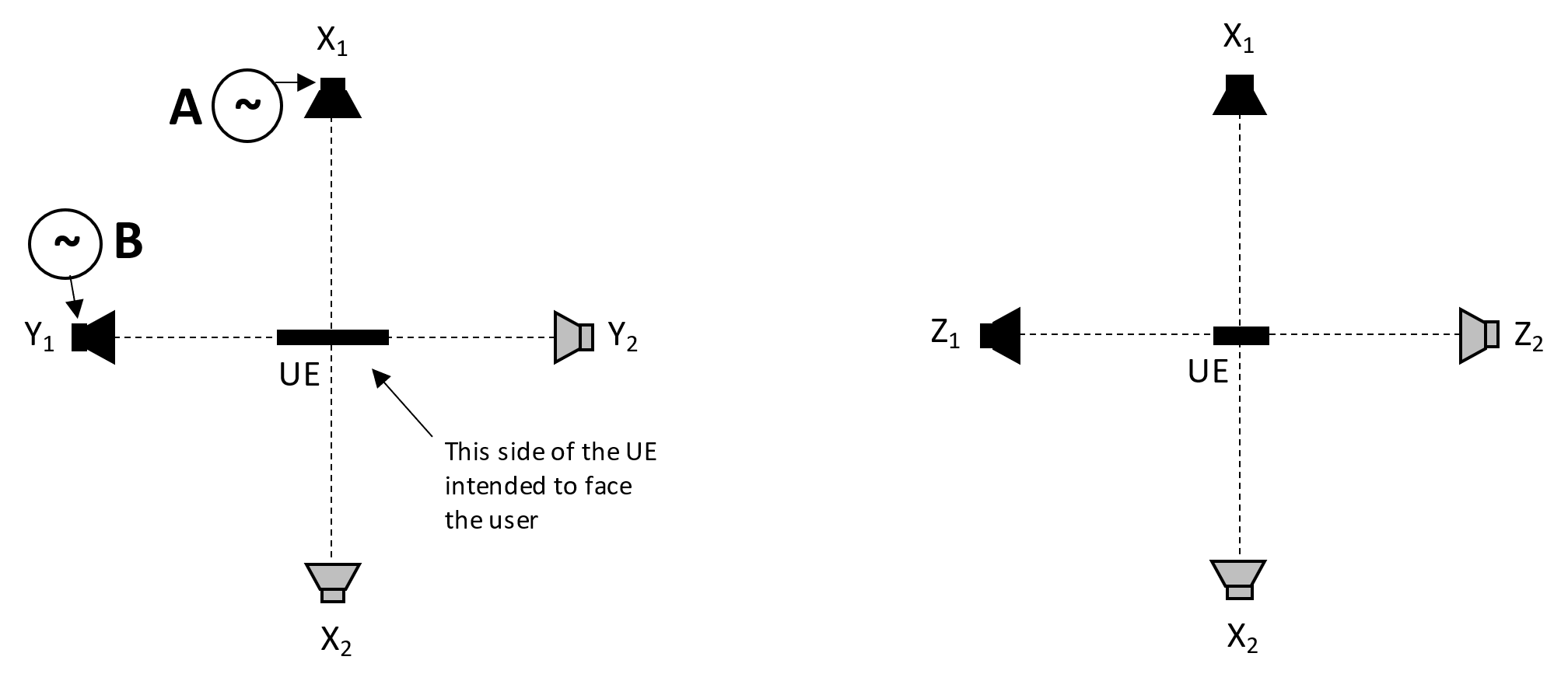


Figure X: Location of loudspeakers

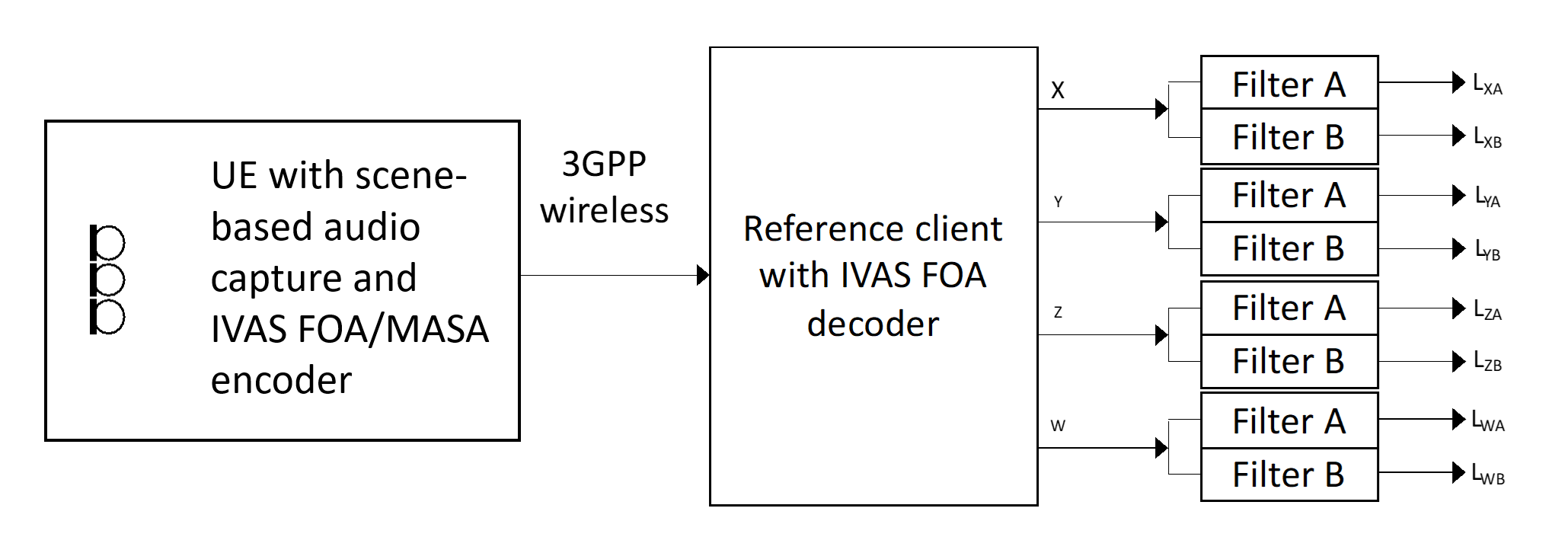


Figure X: Example using FOA; The UE under test is connected to a test system composed of a 3GPP wireless system simulator and a reference client with B-format output and frequency-domain filters for analysis. For HOA, the same setup is used and the higher order ambisonics components at the receiver are ignored. For MASA input capture, IVAS MASA encoder is utilized

[Editor’s note: The actual subjective performance of the immersive capture after rendering using a high-quality reference should be considered. This applies to most test cases with unrealistic test signals and should be addressed in a general section.

### Measurement

The following procedure shall be used:

1. The UE under test is connected to a test system composed of a 3GPP wireless system simulator and reference client with an IVAS session established with B-format output. The codec shall be operated with scene-based audio or metadata-assisted spatial audio input format at [FFS] kbit/s. The audio input format and bitrate shall be reported. The decoder/renderer option shall be FOA.

b) A modulated multi-tone test signal A is played over a loudspeaker at position X1. Simultanously, a modulated multi-tone test signal B is played over a loudspeaker at position Y1. See Annex X for a description of the multi-tone signals.

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

c) The output of each ambisonics component (W, X, Y, Z) is captured. After an initial conditioning time of [5] seconds the remainder of the captured signal is converted to the frequency domain as described in Annex X. The signals are filtered by two different comb filters, filter A and filter B, with passbands corresponding to frequencies in signals A and B respectively. The filters are realized by including/excluding certain frequency bins as described in Annex X.

d) The levels after the filters, averaged over the whole duration, are calculated according to the Annex A.4.

e) The level metrics according to Table X are calculated.

Table X: Assessment of spatial separation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Simultaneous sources | | Requirements on the B-format outputs of the reference decoder | | |
| Source A | Source B | Signal component A | Signal component B | Motivation |
| Position X1 | Position Y1 | LXA – LYA > [N] dB,  |LWA – LXA| < [M] dB | LYB – LXB > [N] dB,  |LWB – LYB| < [M] dB | Signal component A is ideally only seen in X and W, Signal component B is ideally only seen in Y and W |
| |LWA – LWB| < [P] dB | | Signal component A in W equally strong as B in W |
| Position X1 | Position Z1 | LXA – LZA > [N] dB,  |LWA – LXA| < [M] dB | LZB – LXB > [N] dB,  |LWB – LZB| < [M] dB | Signal component A is ideally only seen in X and W, Signal component B is ideally only seen in Z and W |
| |LWA – LWB| < [P] dB | | Signal component A in W equally strong as B in W |
| Position X2 | Position Y2 | LXA – LYA > [N] dB,  |LWA – LXA| < N dB | LYB – LXB > [N] dB,  |LWB – LYB| < N dB | Signal component A is ideally only seen in X and W, Signal component B is ideally only seen in Y and W |
| |LWA – LWB| < [P] dB | | Signal component A in W equally strong as B in W |
| Position X2 | Position Z2 | LXA – LZA > [N] dB,  |LWA – LXA| < [M] dB | LZB – LXB > [N] dB,  |LWB – LZB| < [M] dB | Signal component A is ideally only seen in X and W, Signal component B is ideally only seen in Z and W |
| |LWA – LWB| < [P] dB | | Signal component A in W equally strong as B in W |
| The test is repeated where signals A B are interchanged, to avoid a potential bias. The results from the two tests are averaged. The values M (maximum of difference to omni component), N (minimum of off-axis rejection) and P (maximum unbalance of omnidirectional capture) are TBD. | | | | |

[Editor’s note: in case there will be different specifications for the test methods and the requirements, the table can be moved to the latter document, e.g. TS 26.261.]

[Editor’s note: ground-truth for this test to be clarified]

### Annex A: Test signal and analysis for spatial separation with simultaneous acoustic sources

#### A.1 Test signal definition

The test signal shall be generated according to the ITU-P.501 [1] (subclause 7.2.4.1) and as provided in equation A.1.

A screenshot of a computer

Description automatically generated with low confidence

Figure x– Two channel test signal generation for double-talk evaluations  
based on AM-FM signals

n = 1,2,... (A.1)

where

In ITU-T P.501, the following parameters are defined in a frequency-independent manner: , and . The center frequencies for test signal are defined in the Table A.1.

The frequency-dependent modulation bandwidth is determined as follows:

Table A.1: Centre frequencies and bandwidths (1/3rd octave bands)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Center Frequency [Hz] | Talker | Disable with MASA | Freq. Start [Hz] | Freq. Stop [Hz] | [Hz] |
| 275 | A |  | 269 | 281 | [TBD (should essentially be Fstop – Fstart)] |
| 346 | B | X | 339 | 354 |  |
| 436 | A | X | 427 | 446 |  |
| 550 | B |  | 538 | 562 |  |
| 693 | A | X | 678 | 707 |  |
| 873 | B | X | 855 | 890 |  |
| 1100 | A |  | 1078 | 1123 |  |
| 1386 | B |  | 1358 | 1414 |  |
| 1746 | A |  | 1711 | 1781 |  |
| 2200 | B |  | 2159 | 2241 |  |
| 2772 | A |  | 2731 | 2813 |  |
| 3492 | B |  | 3451 | 3533 |  |
| 4400 | A |  | 4359 | 4441 |  |
| 5543 | B |  | 5502 | 5585 |  |
| 6984 | A |  | 6943 | 7026 |  |
| 8800 | B |  | 8759 | 8841 |  |
| 11087 | A |  | 11047 | 11129 |  |
| 13969 | B |  | 13928 | 14011 |  |

[Editor’s note: Certain frequency components need to be disabled for MASA capture due to characteristics of that format. Whether the frequency components should be disabled in the measurement signal or at the analysis stage is under investigation.]



#### A.2 Shaping filter

Considering the spectra of programme material in general, and considering signal-to-noise ratio in the measurement, the frequency-dependent amplitude of each AM-FM-component is given by the decay d = 3 dB per octave.

The shaping filter is illustrated in Figure A.2.

[TBD]

Figure A.2: Shaping filters for test signal

[Editor’s note: different shaping filters can be considered for general audio and communication scenarios]

#### A.3 Spectral mask

The spectral masks for the calculation of individual per-source / per-talker levels are defined as follows.

The signals are sampled at 48kHz sampling rate and transferred to the frequency domain using a [2^16] FFT, Hann window, [50%] overlap. Frequency bins are multiplied by 1 if they are within the passbands, and by 0 if they are outside.

The passbands of the masks are defined by the stimulus carrier frequencies and the frequency modulation plus a further widening by one frequency bin at each side, see Table A.1.

where

#### A.4 Level calculations

The assessment of the signal separation of two simultaneously captured signals is done by calculating the mean level difference of captured frequency components between assessed channels. The mean level difference is calculated as follows:

1. The level differences of captured frequency component pairs are calculated. Level of each captured frequency component is calculated individually over the mask width for assessed channels:

where and are the power spectrums of the X and Y channels, is the th center frequency and is the th bandwidth of the corresponding test signal component including the modulation bandwidth and frequency mask width.

1. The level difference between two channels’ th frequency components are calculated:
2. Overall spatial separation is evaluated by calculating the average over all the frequency component pair level differences. The mean level difference *N* of the assessed channels is obtained by averaging level difference pairs over frequency component:

[Editor’s note: Channels X and Y are presented here as an example. Same calculations are done between all the assessed output channels with sufficient set of center frequencies, i.e., included center frequencies of the A and B signals.]

]

[

The following method(s) have been incorporated from [5]:

]

[

The following method(s) have been incorporated from [6]:

## Spatial separation for multiple acoustic sources based on multichannel output

**[**

### Test conditions

**Free-field propagation conditions**

- The test environment shall contain a *free-field volume*, wherein free-field sound propagation conditions shall be observed.

- The free-field sound propagation conditions shall be observed down to a frequency of [200Hz].

**[Test environment noise floor]**

[Editor’s note: The test environment noise floor may not have to specified in this clause. Likely, a general clause for the whole specification is sufficient.]

**Loudspeaker array**

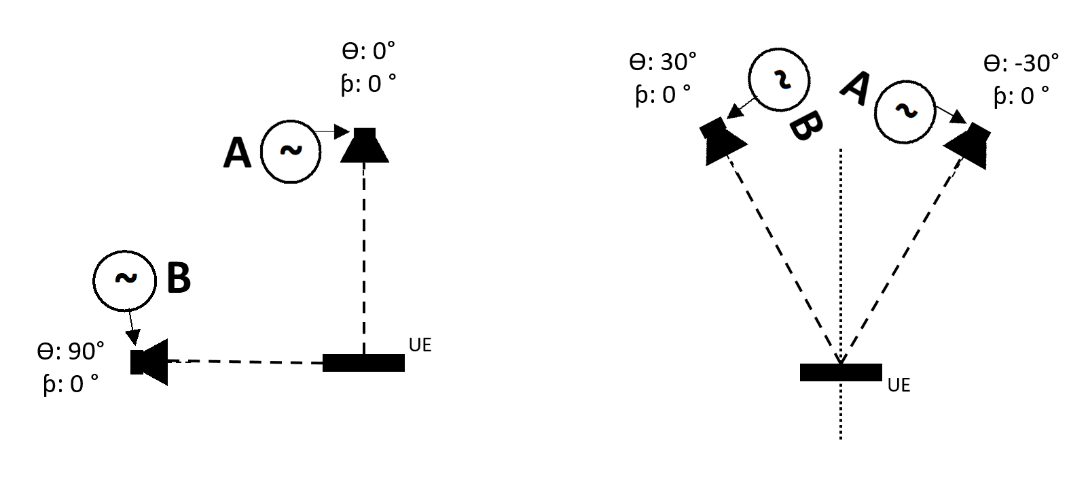
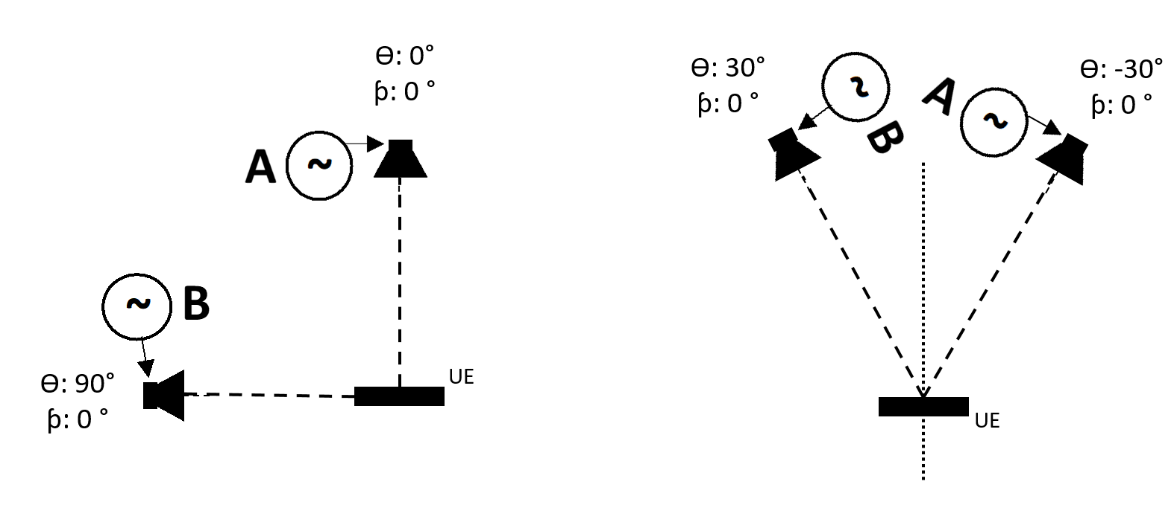
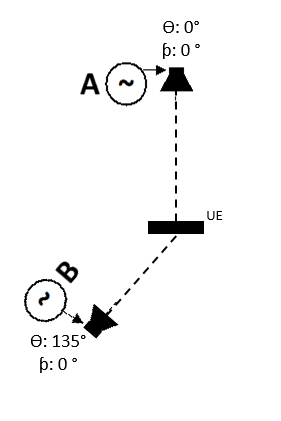
An array of coaxial loudspeakers is located at a set of predefined directions *(qi, fi)*, *i = 1, …,8*, from the geometric center of the UE. The different locations may be realized using multiple loudspeakers or by rotation of at least two loudspeakers or by rotation of the UE.

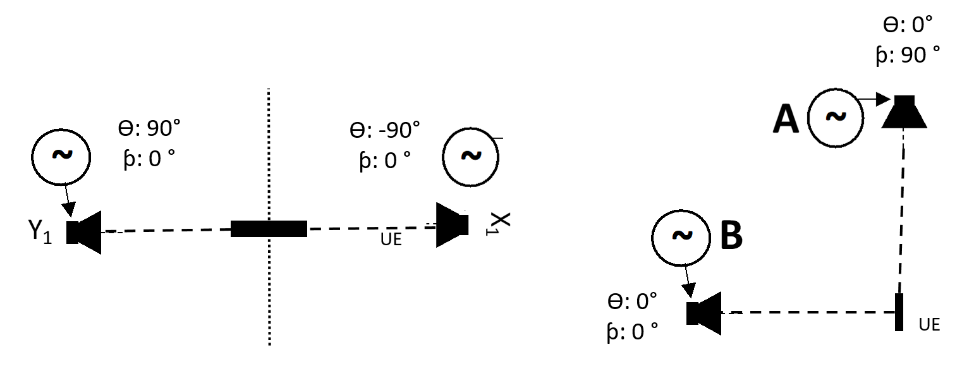
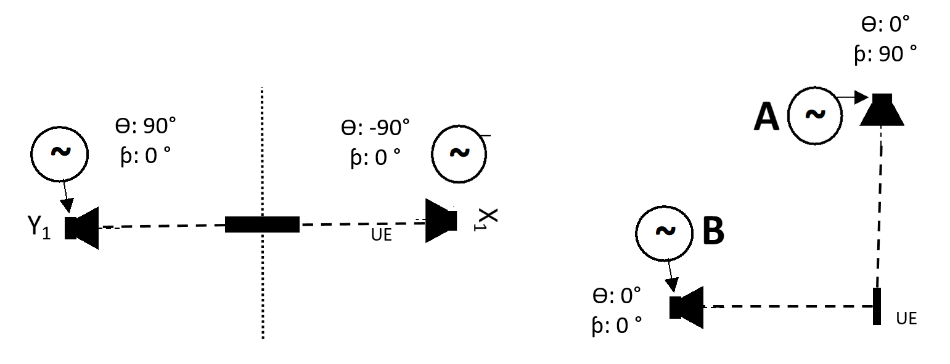
The distance from the loudspeaker front baffle to the center of the UE shall be at least [1m] and equal within ±[x]% for all loudspeakers.

The loudspeakers shall be equalized to a flat frequency response and equal sensitivity with the UE absent, using a [measurement microphone and diffuse-field equalization] placed at the UE position. The microphone shall point in the positive Z direction with its membrane in the XY plane.

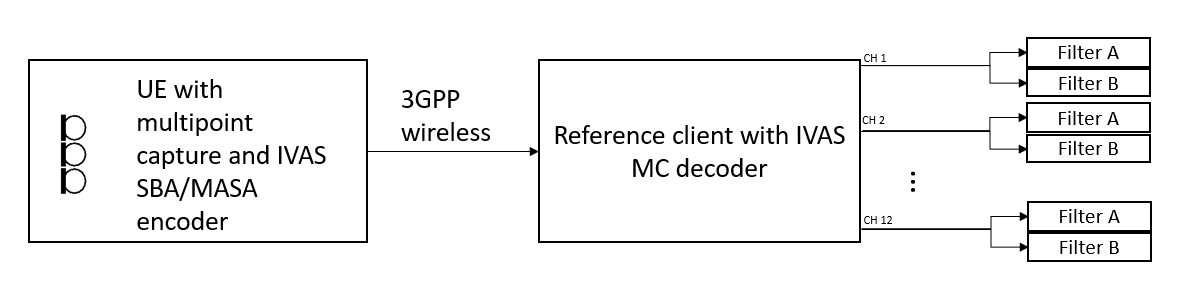
**Table X: Location of loudspeakers generating stimuli in the acoustic chamber, and configuration for the reference client renderer**

|  |  |  |  |
| --- | --- | --- | --- |
| **i** | **qi** [deg] | **fi** [deg] | **Comment** |
| 1 | 0 | 30 | Left, MC channel 1 |
| 2 | 0 | -30 | Right, MC channel 2 |
| 3 | 0 | 0 | Center, MC channel 3 |
| 4 | 0 | 135 | Left Surround, MC channel 5 |
| 5 | 0 | -135 | Right Surround, MC channel 6 |
| 6 | 0 | 90 | Left Side Surround, MC channel 7 |
| 7 | 0 | -90 | Right Side Surround, MC channel 8 |
| 8 | 90 | 0 | One height speaker is used in the acoustic test chamber.   The reference renderer in the reference client is configured with four height speaker signals in accordance with the IVAS default 7.1.4 configuration [reference to future IVAS specification], MC channels 9-12. These four signals are added in the time domain before further measurements. |

**  **

** **

**Figure X: Utilized loudspeaker position combinations**



**Figure X: Example using IVAS SBA or MASA; The UE under test is connected to a test system composed of a 3GPP wireless system simulator and a reference client with 7.1+4 multichannel output and frequency-domain filters for analysis.**

### Measurement

The following procedure shall be used:

1. The UE under test is connected to a test system composed of a 3GPP wireless system simulator and reference client with an IVAS session established with 7.1+4 multichannel output. The encoder shall be operated with scene-based audio or metadata-assisted spatial audio input format at [512] kbit/s. The audio input format and bitrate shall be reported. The decoder/renderer option shall be 7.1+4 multichannel.

b) A modulated multi-tone test signal A is played over a loudspeaker at position iA. Simultaneously, a modulated multi-tone test signal B is played over a loudspeaker at position iB. See Annex A.1 for a description of the multi-tone signals.

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

c) The output of each multichannel output channel (1, 2, …,12) is captured. After an initial conditioning time of [5] seconds the remainder of the captured signal is converted to the frequency domain as described in Annex X. The signals are filtered by two different comb filters, filter A and filter B, with passbands corresponding to frequencies in signals A and B respectively. The filters are realized by including/excluding certain frequency bins as described in Annex A.3.

d) The levels after the filters, averaged over the whole duration, are calculated by summing the power of the selected bins.

e) The level metrics according to Table X are calculated.

Table X Requirements on spatial separation of decoded multichannel output channels. Subscripts denotes the channel number of MC output and applied filtering.

|  |  |  |  |
| --- | --- | --- | --- |
| Simultaneous sources | | Requirements on the 7.1+4 Multichannel outputs of the reference decoder | |
| Source A | Source B | Signal component A | Signal component B |
| Azi: 0°  Ele: 0° | Azi: 90°  Ele: 0° | L3A – L7A > [NA] dB | L7B – L3B > [NB] dB |
| Azi: -30°  Ele: 0° | Azi: 30°  Ele: 0° | L2A – L1A > [NA] dB | L2B – L1B > [NB] dB |
| Azi: -90°  Ele: 0° | Azi: 90°  Ele: 0° | L8A – L7A > [NA] dB | L7B – L8B > [NB] dB |
| Azi: 0°  Ele: 0° | Azi: 135°  Ele 0° | L3A – L5A > [NA] dB | L5B – L3B > [NB] dB |
| Azi: 0°  Ele: 0° | Azi: 0°  Ele: 90° | L3A –LƩ9…12A > [NA] dB | L Ʃ9…12B - L3B > [NB] dB |

Editor’s note: It is expected that the threshold numbers will be set based on experiments with high-quality equipment in free-field propagation conditions with and without IVAS coding.

**]**

[

The following method(s) have been incorporated from [5]:

## Spatial perception test for stereo UE in ATIAS

### 

### Test Conditions

**Free-field propagation conditions**

- The test environment shall contain a free-field volume, wherein free-field sound propagation conditions shall be observed.

- The free-field sound propagation conditions shall be observed down to a frequency of 200Hz

### Test Configurations

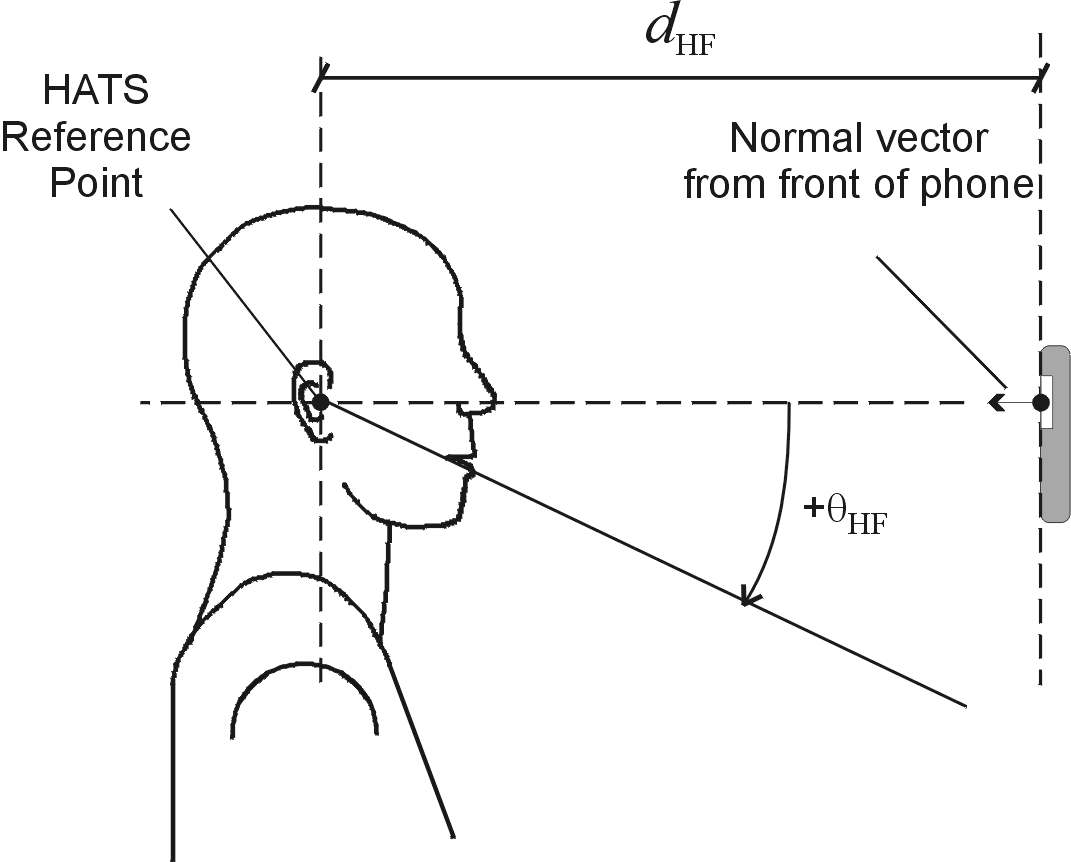
#### Setup for terminals

The setup is referred to TS 26.260 and TS 26.132[2-3]. Including the POI, reference point, etc.

Where the manufacturer gives conditions of use, these will apply for testing. If the manufacturer gives no other requirement, the DUT will be positioned according the reference usage of hand-held hands-free UE in TS 26.132 describing in the following block:

##########################################################################################

If HATS measurement equipment is used, it shall be configured to the hand-held hands-free UE according to figure 4. The HATS should be positioned so that the HATS Reference Point is at a distance dHF from the centre point of the visual display of the Mobile Station. The distance dHF is specified by the manufacturer. A vertical angle HF may be specified by the manufacturer. Where it is not specified, the nominal distance dHF shall be 42 cm and HF shall be 0º.



*Figure 4: Configuration of hand-held hands-free UE relative to the HATS*

################################################################################################

Measurement points:

Diagram

Description automatically generated

**Figure 1: Audio capture block diagram for sending direction measurements**

Editor’s Note: The test should represent what sound the user will get. Hence, the test operator doesn't need to calibrate the DUT. The result should include all the deviations between components in one device (like the sensitivity difference between a microphone array used in DUT) and deviations between different manufactured batches.

#### Measurement method

1. The UE device under test is mounted in the free-field volume such that its reference point is on the axis of the sound source.

Repeat steps b-c) with an azimuth angular resolution of N degrees for every possible usage range (at least cover the visual range):

1. The sound source pointed directly toward the reference point of the DUT, measuring the impulse response of DUT on the α degree from the reference line (minus for left).
2. Change the angle between sound source and DUT.

**Delay Measurement Methodologies**

Refer to TS 26.132 clause 7.10.

**Calculate interchannel time difference and interchannel level difference:**

]

[

## Sending side audio performance assessment for Immersive Audio Systems in wind noise

The following methods have been incorporated from [2]:

### Introduction

This test is applicable to UEs capturing immersive audio, including scene-based (e.g. First and Higher Order Ambisonics), binaural, channel-based (e.g. 7.1.4, 5.1, stereo), and object-based audio.

#### test conditions

- The test conditions should follow the Free-field propagation conditions and test environment noise floor described in TS 26.260[2-1].

- wind speed should be 0m/s.

- The size of free-field volume should be large enough to avoid influencing the wind.

**Wind-generator:**

ETSI TS 103 640[2-2] Annex A lists several turbulent wind generation considerations. Some most important requirements are listed here.

- The acoustic noise should be [TBD]dB less than the wind noise at effective frequency band.

- The airflow wide enough to cover the acoustic test equipment and DUT

- The device must keep the target wind speed stable during the test.

NOTE: this test method is used to measure the overload, the acoustic noise requirement of the device used to generate wind needn’t be so strictly as the requirement in ETSI TS 103 640[2-2] and IEC 60268-4[2-6].

#### Setup for terminals

The setup is referred to TS 26.260[2-1] and TS 26.132[2-3]. including the POI, reference point, etc.

Reference point:

Scene-based: geometric centre. [2-1]

Binaural: centre of the acoustic test equipment EEP-to-EEP axis.[2-2]

Object-based: geometric centre of all transducers.

Multichannel: geometric centre of all transducers.

Position:

When using handset UE, headset or hand-free terminal, the terminal should be placed on HATS, according to the corresponding standard or recommended position.

Handsets are given in ITU-T Recommendation P.64 Annex E.[2-5]

Headsets are given in ITU-T Recommendation P.340[2-4]

Measurement points [2-1]:

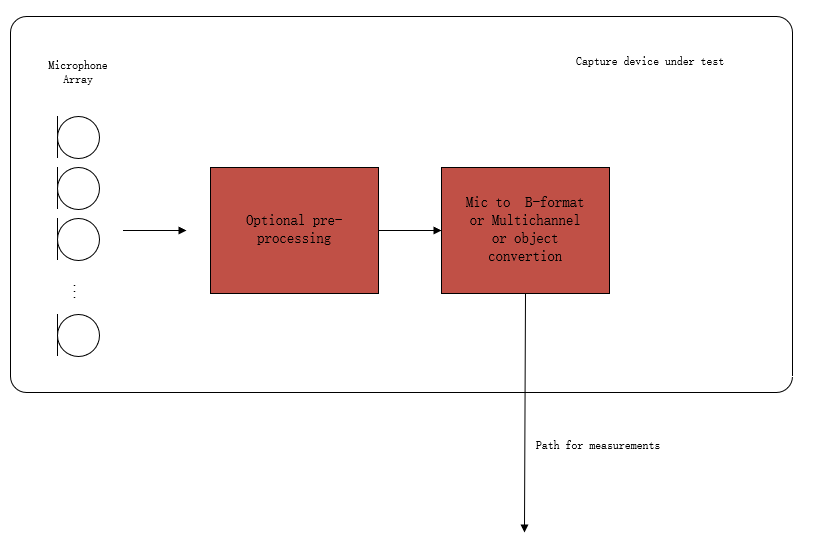


Figure 1: Audio capture block diagram for sending direction measurements

NOTE: The overload point wind speed is a limiting characteristic like the overload sound pressure. All the channels won't affect each other Some processing may cause overload at some special condition, it has damage to communication and is inevitable in windy sensorics, hence the overload caused by processing is included in the result, so select the standard audio signal to measure, and each channel should be measured independently.

#### Definition

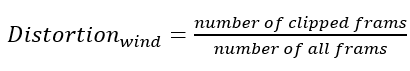
**Overload point wind speed:**

The maximum wind speed at which the distortion of the terminal does not exceed a specified limit(the value of the limit is TBD) for any possible direction of wind incidence and any channel the device outputted.

**Distortion rate:**

Since the clipping appears more frequently at higher wind speed, the probability of clipping appearing in the test signal can indicate distortion rate caused by wind.

So, the source suggests using the rate of clipped frames in all test frames as the distortion rate.



**clipped frames:**

The clipped frames will have any of the following characterises:

1. reach the up limit of signal level (the value needs to test for the DUT)
2. frequency range in high frequency is different from wind noise without clipped.

#### Wind noise measurement method with wind generator for sending direction.

1. The UE device under test is mounted in the free-field volume such that its reference point is on the axis of the wind generator exit port and 30 cm from the exit port.

Repeat steps b-c) with an azimuth angular resolution of N degrees for every possible wind direction:

NOTE 2: Since limiting the wind direction in real usage scenarios is not suitable, the test should be implemented in every possible wind direction.

1. The wind generator is the target wind speed on the DUT, and the airflow should cover the DUT.
2. The output of the UE device is stored for offline analysis. The signal should be stored before the wind start and its duration time should larger than 60s.

Increasing the wind speed and repeating the test until the output signal is overloaded or reaches the expected wind speed.

NOTE 3: the wind speed should be selected carefully to avoid the overload damage caused by wind influence on the later test

**Calculation of wind-resistant ability**

The wind-resistant ability represents with the wind speed overload point, which means the terminal can work stable in all directions, and all channels with the wind speed don't exceed the overload point.

The terminal, with several audio channels output, should be calculated by every channel.

The following conclusion has been incorporated from [3]:

### Recommendations for wind noise simulations for terminal testing

* Wind noise simulations for terminal testing have to be carefully defined under the following constraints:
  + A minimum degree of laminar flow should be ensured by means of e.g., spatial wind speed accuracy, measured at multiple points.
  + A certain degree of reproducibility should be ensured across labs and/or different test equipment solutions.
  + The noise produced by the ventilator/generator should not exceed a certain threshold to minimize the impact on the actual measurements.
  + For employment in typical measurement rooms, a manageable generator size is required – which might limit the aforementioned constraints even further.
* There is currently no specification available or known to the group that uses or defines such a wind noise simulation.
* Possible test methods and performance requirements for ATIAS should be limited to certain form factors/types of terminals. Wind noise simulation for smaller devices is most likely more feasible and reproducible than for larger ones.
* If applicable, specification of a wind noise simulation, test methods and performance requirements should be verified by round robin tests.

]

# Candidate receiving side test methods and requirements

[

TBD.

]

# References

[1] S4-221449: On send side audio performance assessment for Immersive Audio Systems –additional metrics, Dolby Laboratories Inc.

[1-1] S4-220482: On ATIAS acoustic performance testing for FOA audio, Dolby Laboratories Inc.

[1-2] S4-191167: Description of the IVAS MASA C Reference Software, Nokia Corporation

[1-3] GIF image: <https://commons.wikimedia.org/wiki/Category:Microphone_polar_patterns>

[2] S4-221353: Proposal of wind noise test in ATIAS, Beijing Xiaomi Mobile Software

[2-1] 3GPP TS 26.260: " Objective test methodologies for the evaluation of immersive audio systems."

[2-2] ETSI TS 103 640 V1.2.1-Test Methods and Performance Requirements for Active Noise Cancellation Headsets and other Earphones

[2-3] 3GPP TS 26.132: " Speech and video telephony terminal acoustic test specification."

[2-4] ITU-T Recommendation P.340 (05/2000): "Transmission characteristics and speech quality parameters of hands-free terminals".

[2-5] ITU-T Recommendation P.64 (06/2019): "Determination of sensitivity/frequency characteristics of local telephone systems".

[2-6] IEC 60268-4: Sound system equipment - Part 4: Microphones

[3] S4-230035: Wind noise generation for terminals, HEAD acoustics

[4] S4-230259: Spatial audio capture – spatial separation for multiple acoustic sources based on FOA components, Dolby Laboratories Inc., Nokia Corporation, HEAD acoustics

[5] S4-230189: Add the spatial perception test for stereo UE in ATIAS, Xiaomi

[6] S4-230231: On spatial separation for multiple acoustic sources based on multichannel output, Nokia Corporation

[6-1] Recommendation ITU-T P.501, Test signals for use in telephony and other speech-based applications

[6-2] Tdoc S4-221297: IVAS Design Constraints (IVAS-4)

[7] S4-230232: On direction-of-arrival estimation for MASA input format, Nokia Corporation

[7-1] Tdoc S4-221449: On send side audio performance assessment for Immersive Audio Systems –additional metrics

[7-2] Tdoc S4-221297: IVAS Design Constraints (IVAS-4)

# Revision history

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Meeting** | **Subject/Comment** | **Old** | **New** |
| 2022-11-17 | SA4#121 | Initial version incorporating S4-221449 and S4-221353 | N/A | 0.1.0 |
| 2023-02-24 | SA4#122 | Agreed updates from SA4#122 | 0.1.0 | 0.2.0 |
| 2023-03-23 | Audio SWG post SA4#122 | Editorial updates | 0.2.0 | 0.2.1 |
| 2023-03-24 | Audio SWG post SA4#122 | Editorial updates | 0.2.1 | 0.2.2 |
| 2023-04-19 | SA4#123-e | Incorporation of Tdocs 543, 544, 545 | 0.2.2 | 0.3.0 |
| 2023-05-25 | SA4#124 | Consolidations of comments, updates based on Tdocs 910 and 911, re-arrangement of text in section 4.7 | 0.3.0 | 0.4.0 |

1. Stefan Bruhn, e-mail: stefan.bruhn@dolby.com [↑](#footnote-ref-1)
2. Depending on the defined format for the ambisonics, scaling factors may be applied to the X and Y components in this example [↑](#footnote-ref-2)