**3GPP TSG-SA4 Meeting SA4#111-e  *S4-201612***

**, , - revision of S4-201309**

|  |
| --- |
| *CR-Form-v12.0* |
| **CHANGE REQUEST** |
|  |
|  | **26.260** | **CR** | **0000** | **rev** | **--** | **Current version:** | **16.2.0** |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

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

|  |
| --- |
|  |
| ***Title:***  | Immersive Speech Communication Systems |
|  |  |
| ***Source to WG:*** | HEAD acoustics GmbH |
| ***Source to TSG:*** | S4 |
|  |  |
| ***Work item code:*** | ATIAS |  | ***Date:*** | 2020-11-20 |
|  |  |  |  |  |
| ***Category:*** | **B** |  | ***Release:*** | Rel-17 |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)Rel-12 (Release 12)**Rel-13 (Release 13)Rel-14 (Release 14)Rel-15 (Release 15)Rel-16 (Release 16)* |
|  |  |
| ***Reason for change:*** | The objective of the work item ATIAS is to develop a set of test specifications for 3GPP immersive services, both conversational and non-conversational. The current version of TS 26.260 does not contain any tests methodologies that are specifically tailored to conversational services. |
|  |  |
| ***Summary of change:*** | Introduction of new clauses for test setups for immersive communication systems. |
|  |  |
| ***Consequences if not approved:*** | Testing of immersive conversational services not considered. Evaluation of such a service not possible. |
|  |  |
| ***Clauses affected:*** | 2, 5.1 (new), 5.2 (new), 5.3 (new), 5.4 (new) |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **x** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ... |
| ***(show related CRs)*** |  | **x** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

------------------------- START OF CHANGE 1 -------------------------

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[…]

[

[8] ITU-T Recommendation P.341: "Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals".

[9] ITU-T Recommendation P.501: "Test signals for use in telephony and other speech-based applications".

[10] ITU-T Recommendation P.56: "Objective measurement of active speech level".

[11] ITU-T Recommendation P.57: "Artificial ears".

[12] ITU-T Recommendation P.58: "Head and torso simulator for telephonometry".

]

------------------------- END OF CHANGE 1 -------------------------

------------------------- START OF CHANGE 2 -------------------------

# 5 Objective Test Methodologies for Immersive Conversational Systems

## [5.1 Interfaces]

For the testing of immersive communication systems, the a reference client as shown in Figure 1 is required. The application of ambient sound/noise is separately specified for each test case, where applicable.



Figure 1: Interfaces for specification of terminal acoustic/audio characteristics.

The following definitions apply:

- Pose: Orientation of the listener or device. This can be used when e.g., the audio processing (such as head-related transfer function) is modified based on a head tracker. Spherical coordinates according to the definition in clause 3.1 can be used to specify the pose.

- Acoustic reference points (of head and torso simulator representing the user of the UE):

- ERP/DRP…: ear reference point or eardrum reference point (with or without diffuse-field correction)

- MRP: mouth reference point

- Electrical reference points:

- ELR: electrical receive measurement point of the UE. This may be analogue or digital, it may consist of multiple channels.

- ELS: electrical send measurement point of the UE. This may be analogue or digital, it may consist of multiple channels.

- RREF: receive reference point of the test system. This may be analogue or digital, it may consist of multiple channels. In case of conversational, it corresponds to the POI (point of interconnect).

- SREF: send reference point of the test system. This may be analogue or digital, it may consist of multiple channels. In case of conversational, it corresponds to the POI (point of interconnect).

Definitions also copied from S4-200125 and may be aligned with definitions/naming conventions from IVAS codec.

## [5.2 Reference and Test Setups for Assessment of Immersive Conversational Systems]

### 5.2.1 Introduction

[Testing of immersive communication systems is a more challenging task than testing conventional communication systems. While all the important aspects of conventional systems are still of interest, there are additional degrees of freedom that an immersive system offers which need to be considered in the test design and accordingly in the test setup.

Dependency on IVAS; different supported audio formats should be mentioned/considered here.

The realistic assessment of such conversational systems needs to relate to the perception of the users on the receiving end. Accordingly, testing of the sending direction requires a reference receving client according to clause 5.1. The combination of a sending DUT with the receiving client allows for a comparison of the immersive conversational system with the predefined reference setups.

Testing of the receiving direction requires a reference sending client according to clause 5.1, including playback of prerecorded signals via the point of interconnect. The combination of a sending client with the receiving DUT allows for a comparison of the immersive conversational system with the predefined reference setups.]

### [5.2.2 Round-table Conference]

First proposal, further investigation and initial measurements are needed.

#### 5.2.2.1 Reference Setup

The reference setup for the conferencing service consists of multiple talking and listening head and torso simulators (HATS, according to ITU-T P.58 [12]) in a static arrangement. Figure 1 illustrates the situation: a circular distribution with three talking HATS and three listening HATS located around a table. The table has a diameter of 100 cm and the HATS are positioned in the group audio terminal position from ITU-T P.341 [8] (i.e. at a distance of 80 cm from the center of the table facing the center of the table). The three positions on the talking side of the table are denoted A1, A2 and A3 in the following. The three positions on the listening side are B1, B2 and B3. The angles between neighboring positions on the talking as well as on the listening side are 45°.



Figure 1: Example communication scenario

A logarithmic sweep is measured from each talker (MRP) to each listener (DRP with DF-equalization) with a nominal source level of -4.7 dB Pa at each talker. To increase SNR for the sweep measurements, a gain of at least +6 dB (according to clause 6.2.3 of ITU-T P.58 [12]) is recommended.

TBD: define sweep/IR parametrization

Impulse responses and transfer functions are determined from the sweep recordings. The following metrics / curves shall be calculated for all transmission paths based on these transfer characteristics:

* Interaural time differences (ITDRef)
* Binaural frequency response (BFRRef)
* [TBD, other metrics]

[Maybe it makes sense to swap clauses 5.1 and 5.2, so that metric description can be referenced here?]

#### 5.2.2.2 Test Setup

For the measurement according to the round-table conference setup, table and at up to six HATS are arranged as described in clause 5.1.2.1. If applicable, the DUT is positioned at the center of the table.

### [5.2.3 Another reference scenario]

[Clause 5.1.3 (and following) and may describe more/different/other reference scenarios]

## [5.3 Metrics for Assessment of Immersive Conversational Systems]

Tentiative metrics; usage of metrics depend on the agreed tests.

### 5.3.1 Introduction

This clause introduces several metrics, which are commonly used in the assessment of immersive conversational systems. If necessary, distinctions between sending and receiving direction are specified.

Most metrics rely on the analysis of a binaural signal as perceived by the user of the conversational system – independent of the actual configuration of the system (like e.g., presentation via binaural rendering or playback via loudspeakers). In these cases, the signals at the user positions shall be recording with a HATS with DF-equalization according to ITU-T P.58 [12], which is equipped with artificial ears of type 3.3 according to ITU‑T P.57 [11]. The binaural signals recorded in this way are denoted as for left and for right ear in the following.

### 5.3.2 Interaural time differences

#### 5.3.2.1 Introduction

It is known from studies on human perception that interaural differences in level or time are among the strongest features that are used to localize sound sources. The interaural time difference can be determined robustly, e.g., from the lag of the maximum of the cross correlation between the left and right channel.

The interaural level difference can be seen, e.g., in the frequency responses for the transmission paths. Due to the frequency dependent impact of head shadowing, larger differences can be observed for higher frequencies.

#### 5.3.2.2 Definition

The interaural time difference can be calculated by locating the maximum in the cross correlation between the recordings and dividing the resulting lag by the sampling frequency :

### 5.3.3 Binaural frequency responses

#### 5.3.3.1 Introduction

Frequency responses are a typical methodology to characterize the transmission behaviour of a conversational system. Analyzing the binaural frequency response reveals perceptually relevant properties of the system.

#### 5.3.3.2 Definition

The binaural frequency responses are determined individually for left and right ear. The recorded signals are referred to the test signal x(k) in the frequency domain:

Measurements shall be made at [1/12]-octave intervals as given by the [R.40] series of preferred numbers in ISO 3 [xx] for frequencies from [100 Hz to 20 kHz] inclusive. For the calculation of , the averaged measured level at each frequency band is referred to the averaged test signal level measured in each frequency band.

If not specified otherwise, the test signal x(k) shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [9], calibrated to -4.7 dBPa active speech level according to ITU-T P.56 [10] at MRP of the talker (or equivalently calibrated, e.g., at electrical point of interconnect).

### 5.3.4 Interaural level differences

#### 5.3.4.1 Introduction

Level differences between the ears in higher frequency regions are a strong cue for the spatial perception. These level differences can be identified by a comparison of the frequency responses between left and right ear.

#### 5.3.4.2 Definition

The interaural level difference is expressed as the difference in spectrum between left and right ear:

Measurements shall be made at [1/12]-octave intervals as given by the [R.40] series of preferred numbers in ISO 3 [xx] for frequencies from [100 Hz to 20 kHz] inclusive. To determine , two calculation methods are possible:

1) The averaged measured level at each frequency band of left ear is referred to the averaged measured level at each frequency band of the right ear .

2) The frequency response as determined according to clause 5.2.3 is referred to the corresponding response at each frequency band.

Both calculation methods are equivalent.

If not specified otherwise, the test signal x(k) shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [9], calibrated to -4.7 dBPa at MRP of the talker (or equivalently calibrated, e.g., at electrical point of interconnect).

## [5.4 Objective Test Methodologies for Assessment of Immersive Conversational Systems in the Sending Direction]

### 5.4.1 Introduction

For each reference/test setup described in clause 5.1, the following sub-clauses specify test methodologies for the assessment of immersive conversational systems in the sending direction.

### [5.4.2 Round-table Conference]

Depends on the decision if this setup is included or not.

#### 5.4.2.1 Introduction

The DUT is setup according to clause 5.1.2.2.

Talking HATS shall be positioned at A1, A2 and A3 as described in clause 5.1.2.2, in relation to the DUT (if applicable).

In order to obtain binaural signals for analysis, a reference/suitable receiving client shall be connected to the DUT in sending direction. The reference receiving client shall be configured in relation to the listening HATS positioned at B1, B2 and B3 according to clause 5.1.2.2. The reference client might be replaced by defined simulation (e.g., a binaural renderer), if an appropriate point of interconnect is available.

#### 5.4.2.2 Interaural time differences

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [9], calibrated to -4.7 dBPa active speech level according to ITU-T P.56 [10] at MRP.

b) The test signal is played back for each talker position A1, A2 and A3 in sending direction.

c) The measured binaural signals are captured at the (virtual) listening positions B1, B2 and B3 at DRP with DF-equalization.

d) The interaural time difference ITDTest,SND is calculated according to clause 5.2.2 and reported for all paths. For example, the ITD from talker A2 to (virtual) listener B2 is denoted as ITDTest,SND(A2, B2).

e) The difference in ITD between test and reference setup is determined for each talker (A) and listener (B) position as:

#### 5.4.2.3 Binaural frequency responses

[TBD]

#### 5.4.2.4 Interaural level differences

[TBD]

## [5.5 Objective Test Methodologies for Assessment of Immersive Conversational Systems in the Receiving Direction]

### 5.5.1 Introduction

For each reference/test setup described in clause 5.1, the following sub-clauses specify test methodologies for the assessment of immersive conversational systems in the receiving direction.

### [5.5.2 Round-table Conference]

#### 5.5.2.1 Introduction

The DUT is setup according to clause 5.1.2.2.

Binaural signals for analysis are recorded by the listening HATS, which shall be positioned at B1, B2 and B3 as described in clause 5.1.2.2, in relation to the DUT (if applicable).

For the insertion into receiving direction, a reference/suitable sending client shall be connected to the DUT in receiving direction. The reference sending client shall be configured related to the talking HATS positioned at A1, A2 and A3 according to clause 5.1.2.2. If an appropriate point of interconnect is available, an electrical test signal can directly be inserted into receiving direction of the DUT.

#### 5.5.2.2 Interaural differences

a) The test signal to be used for the measurements shall be the British-English single talk sequence described in ITU-T Recommendation P.501 [9]. The test signal shall be calibrated to -4.7 dBPa active speech level according to ITU-T P.56 [10] at MRP of the talker at the sending side, or equivalent (for point of interconnect).
[To do: how to specify electrical multi-channel levels?]

b) The test signal is played back for each (virtual) talker position A1, A2 and A3 in sending direction via the sending reference client or via point of interconnect.

c) The measured binaural signals are captured at the listening positions B1, B2 and B3 at DRP with DF-equalization.

d) The interaural time difference ITDTest,RCV is calculated according to clause 5.2.2 and reported for all paths. For example, the ITD from (virtual) talker A2 to listener B2 is denoted as ITDTest, RCV(A2, B2).

e) The difference in ITD between test and reference setup is determined for each talker (A) and listener (B) position as:

#### 5.5.2.3 Binaural frequency responses

[TBD]

#### 5.5.2.4 Interaural level differences

[TBD]

------------------------- END OF CHANGE 2 -------------------------