ETSI TR 126 996 V18.1.0 (2024-10)

TECHNICAL REPORT

Universal Mobile Telecommunications System (UMTS);
LTE;
5G;
Immersive Audio for Split Rendering Scenarios;
Performance characterization
(3GPP TR 26.996 version 18.1.0 Release 18)

Reference

RTR/TSGS-0426996vi10

Keywords

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under https://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

shall indicates a mandatory requirement to do something

shall not indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

should indicates a recommendation to do something

should not indicates a recommendation not to do something

may indicates permission to do something

need not indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

can indicates that something is possible

cannot indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

will indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

will not indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

might indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

might not indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

is (or any other verb in the indicative mood) indicates a statement of fact

is not (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

Introduction

An essential architectural characteristic of XR clients is the reliance on a functional split between a set of composite pre-renderers that are implemented as parts of a presentation engine and a set of post-rendering operations implemented on an End Device prior to final output. Split rendering may be a necessity if the End Device is power constrained or limited in computational power. However, split rendering is not precluded from other End Devices that do not have such constraints. A discussion of relevant split rendering scenarios is provided in TR 26.865 [2], together with general design guidelines for immersive audio split rendering systems and specific design constraints and performance requirements for split rendering solutions for the 3GPP IVAS codec [3]. The latter are the basis for defining the ISAR baseline solutions as specified in TS 26.249 [4] and as realized in the split rendering feature of the IVAS codec [3],[10],[11].

1 Scope

This TR presents a characterization of the ISAR split rendering solutions specified in TS 26.249 [4] in terms of audio quality performance and technical properties. The characterization is done for the ISAR baseline solution that is defined as a feature of the IVAS codec.

NOTE: As the ISAR split rendering baseline solution is also incorporated as a feature into the IVAS codec, some of the characterization results of the present document may also appear in similar form in the IVAS characterization TR 26.997 [5]. Both representations are equally valid.

2 References

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

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TR 26.865: "Immersive Audio for Split Rendering Scenarios; Requirements ".

[3] 3GPP TS 26.250: "Codec for Immersive Voice and Audio Services (IVAS); General overview".

[4] 3GPP TS 26.249: “Immersive Audio for Split Rendering Scenarios; Detailed Algorithmic Description of Split Rendering Functions“.

[5] 3GPP TR 26.997: “IVAS codec performance characterization”.

[6] 3GPP TS 26.119: “Media Capabilities for Augmented Reality”.

[7] 3GPP TR 26.998: “Support of 5G glass-type Augmented Reality / Mixed Reality (AR/MR) devices”.

[8] 3GPP TR 26.806: “Study on Tethering AR Glasses – Architectures, QoS and Media Aspects”.

[9] 3GPP TR 26.928: “Extended Reality (XR) in 5G”.

[10] 3GPP TS 26.253: "Codec for Immersive Voice and Audio Services (IVAS); Detailed Algorithmic Description including RTP payload format and SDP parameter definitions ".

[11] 3GPP TS 26.258: "Codec for Immersive Voice and Audio Services (IVAS); C code (floating-point)".

[12] Recommendation ITU-R BS.1534 (10/2015): Method for the subjective assessment of intermediate quality level of audio systems.

3 Definitions of terms, symbols and abbreviations

3.1 Terms

Void

3.2 Symbols

Void

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

CLL	Cross-check Listening Laboratory
CuT	Codec under Test
DoF/DOF	Degree of Freedom
FB	Full Band
HOA3	Higher-Order Ambisonics, 3rd order
IVAS	Immersive Voice and Audio Services
ISAR	Immersive Audio for Split Rendering
ISM	Independent Streams with Metadata
LCLD	Low Complexity Low Delay
LC3plus	Low Complexity Communication Codec Plus
LL	Listening Laboratory
MASA	Metadata-Assisted Spatial Audio
MUSHRA	Multi Stimulus test with Hidden Reference and Anchor
PCO	Processing Cross-check Organization
SP	Solution Proponent
SPL	Solution Proponent Laboratory
TC	Transport Channel

4 General

4.1 Project History

The work item "Immersive Audio for Split Rendering Scenarios (ISAR)" started early 2023 with the target to provide split immersive audio rendering solutions within Rel-18 timeframe. Essential milestones of this work were:

- Provisioning of requirements underlying the work in TR 26.865 [2]
- Provisioning of TS 26.249 describing split immersive audio rendering solutions meeting the requirements
- Provisioning of CRs to IVAS codec specifications [10]-[12] adding the ISAR split rendering feature to IVAS
- Provisioning of this TR documenting quality performance and technical properties of the specified split immersive audio rendering solutions.

An important split of the work was to develop IVAS specific solutions ISAR baseline solutions under a work track A and to develop and specify codec/renderer agnostic solutions under a work track B. This split is reflected in both the ISAR TS [4] and this TR where baseline solutions in the context of the IVAS codec and additional ISAR features are described separately.

The selection of IVAS specific solutions to be standardized as ISAR baseline followed a rigorous selection process within 3GPP SA4. The selection process applied the following selection rules:

Rule 1: Provision of full set of selection phase deliverables

The proponents of a candidate solution shall provide all required items listed in PD on ISAR/IVAS selection deliverables in due time in order to be considered further in the selection process of the split rendering solution of the IVAS codec standard.

Rule 2: Compliance with design constraints

The proponents of a candidate solution shall report on compliance of the candidate solution with the IVAS related ISAR design constraints in TR 26.865 [2].

Rule 3: Performance

The performance of the candidate solution(s) will be analysed against the IVAS related ISAR performance requirements in TR 26.865 [2].

The selection procedure involved the following steps:

The selection procedure will consist of the following pre-selection steps for a given candidate solution:

1. The selection deliverables associated with the candidate solution according to rule 1 are reviewed by SA4 and a determination will be made if they meet the requirements.

2. The compliance with design constraints of the candidate solution is evaluated based on the report provided according to rule 2 and a determination will be made if the design constraints are met.

3. The performance of the candidate solution will be analysed against the IVAS related ISAR performance requirements in TR 26.865 [2] and a determination will be made if the performance requirements are met.

4. Based on the outcome of steps 1-3, SA4 will discuss and determine whether the candidate solution is eligible to be adopted as split rendering solution of the IVAS codec standard. This ends the pre-selection steps for the candidate.

The selection procedure will further consist of the following main selection steps for the pre-selected candidate solution(s):

1. If there is only a single pre-selected candidate solution, its status will be elevated to selected candidate solution.

2. If there are more than a single pre-selected candidate solution, SA4 will enter a discussion which of the pre-selected candidate solutions has the highest merit in terms of meeting or exceeding ISAR WID objectives and particularly the relevant IVAS specific performance requirements and design constraints. SA4 will then seek agreement on the most meritful solution and elevate its status to selected candidate solution.

3. Agreement will be declared on the selection.

4. SA will be requested to approve the selection and the relevant associated deliverables such as CRs to IVAS specifications.

The actual selection considered only a single candidate solution.

4.2 Overview of the ISAR Work Item

4.2.1 Work item justification

TS 26.119 [6] assumes a common XR Baseline Client architecture. An essential characteristic is that a functional split is envisioned between a Presentation Engine comprising a set of composite renderers that are controlled by a Scene Manager and an XR Runtime performing a set of functions that interface with a platform to perform commonly required operations, e.g. post-rendering, prior to final output. The relevant interface between Presentation engine and end device may be a 5G physical interface between, e.g., between a smartphone or 5G EDGE and a lightweight device (AR glasses) like those considered in 5G EDGe-Dependent AR (EDGAR) and 5G Wireless Tethered AR UEs as described in TR 26.998 [7] or those considered in TR 26.806 [8].

The functional split assumed in split renderer architectures is a result of stringent implementation and operational requirements applicable for rendering of XR media on XR devices. For head-tracked immersive audio, the need to rely on a split renderer architecture, may depend on various factors among which the round-trip latency between the renderer in the presentation engine and the lightweight device is a decisive parameter. There are scenarios where this latency may be substantial which may prefer a split rendering approach with pose correction in the end device for binaural audio in a similar way as for video unless decoding and head-tracked binaural audio rendering on the lightweight device does not exceed its strict complexity constraints. In other scenarios, that latency may be sufficiently low, in which case the head-tracked binaural rendering can exclusively be done in the presentation engine. It is notable that the transmission over the interface may generally be bit rate constrained and dependent on the specific physical interface.

Binaural audio rendering comprises of signal processing functionalities that may include:

- Binauralization of audio input based on head rotation (3DoF),
- Binauralization of audio input based on listener position and head rotation (6DoF),
- Room acoustics synthesis.

Audio input to be rendered may be a combination of diegetic immersive (3D audio) and non-diegetic sounds. The diegetic immersive sounds need to be binauralized using the up-to-date head rotation data. The head rotation data is typically originating from the head-tracker available from the lightweight end device. The room acoustic synthesis can be performed using room impulse response data or parametric representation thereof, typically supplied to the Presentation Engine.

Depending on constraints and design preferences of the lightweight device (AR glasses, earbuds, etc.) and the properties of the interface between Presentation Engine and end device, solutions are needed that among more are compliant with TR 26.928 [9] and TR 26.998 [7]. The solutions shall address given interface characteristics and not impose any new requirements for them.

Another aspect is the currently ongoing standardization of the EVS Codec Extension for Immersive Voice and Audio Services (IVAS) codec. While low complex rendering for lightweight devices is not a specific design objective, the IVAS codec work item should ideally provide solutions that would enable using IVAS services over head-tracked lightweight clients meeting relevant requirements.

Bearing in mind the evolution of the AR/XR technologies, it would be desirable to design low complex solutions for head-tracked binaural audio rendering on lightweight devices that under certain limitations are agnostic in a sense that the pre-renderer component in the presentation engine could be connected with any immersive binaural audio framework through suitable APIs.

The solutions to be specified are intended to add to the number of rendering options to enable immersive audio services on a broad range of devices, including light-weight AR glasses or earbuds. The pre-rendering part of the solutions is expected to become non-mandatory but shall fulfill the relevant requirements set out under this work item. It should interface through a fully specified intermediate bitstream with a fully specified split rendering decoder. For end device implementations claiming support of a specific solution, a fully compliant implementation of at least the split rendering decoder shall be required. Other end device implementations not claiming support of a specific solution remain at the discretion of the implementor.

4.2.2 Work item objectives

The overall objective of this work item is to develop solutions for immersive binaural audio on head-tracked devices that are compatible with the envisaged split architectures (TS 26.119 [6], 26.998 [7]). The solutions should consider low-complex and lightweight devices and demonstrate operational benefits over solutions with full decoding and rendering in the end device. The following objectives should be achieved with the work item:

- Provide format specification for intermediate representation(s).
    - Provide functional requirements for (pre-)renderer operations to be carried out by Presentation Engine.
    - Define suitable APIs.
- Provide encoder, bitstream and decoder specification for intermediate representations including audio with and without post-rendering control metadata.
- Provide a specification for decoded intermediate representations to provide binaural audio output with and without head-tracker input and post-rendering control metadata.
- Consider potential solutions offered by the IVAS work item, and specify the necessary interfaces.

The work item shall in a first phase identify and agree relevant requirements to be documented in a TR. This shall cover:

- Design constraints related to complexity and memory as well as constraints related to relevant interfaces between presentation engine and end device such as bit rate, latency, down- and upstream traffic characteristics.

- Design constraints related to functional capability requirements such as rendering of non-diegetic sounds, 3DoF rendering of diegetic immersive sounds, 6DoF rendering of diegetic immersive sounds, including simultaneous rendering of different sound categories, and room acoustics synthesis.

- Performance requirements.

The solution(s) are characterized for the range of relevant interface characteristics between presentation engine and lightweight device. The case where the immersive audio is decoded and rendered within the end device should be considered as a reference.

The requirements will be documented in a first technical report. The developments under this work item shall lead to a new specification defining among others textual descriptions of the involved renderers and codec (incl. frame loss concealment) of the intermediate representation(s). The performance of the developed solutions in relation to the requirements will be documented in a second technical report. Solutions meeting the ISAR split rendering requirement may be added to the set of IVAS codec specifications (by means of CRs) if they are found suitable for IVAS. The developed solutions should also be referenced in TS 26.119 [6].

Specific split rendering solutions for IVAS should comprise a non-mandatory default split rendering encoder for the specified internal and stand-alone IVAS renderers. In addition, for a given specific solution there should be specified interfaces offering the possibility either to connect a given (proprietary) renderer for IVAS to the intermediate representation encoder or to use proprietary pre-renderers/intermediate encoders to produce compliant intermediate bitstreams. Such proprietary solutions shall be compliant with the relevant requirements documented in the first technical report. ISAR end device implementations for IVAS claiming support of a specific solution shall be required to have at least a fully compliant split rendering decoder. Other decoder/post-renderer implementations not claiming support of a specific ISAR solution for IVAS remain at the discretion of the implementor.

5 Terms of Reference

The design constraints and performance requirements defined in the requirements TR 26.865 [2] constitute the Terms of Reference for the ISAR work. Notably, specific design constraints and performance requirements could only be defined given IVAS as target codec/renderer of the baseline ISAR solutions. For codec/renderer agnostic solutions, the TR provides guidelines that can be turned into specific design constraints and performance requirements once a specific target system is defined.

6 Audio Quality evaluations

6.1 ISAR Baseline quality

6.1.1 Selection tests

6.1.1.1 Test plan

The selection tests evaluating the performance of the single ISAR candidate solution for IVAS was carried out according to a permanent document on Testing Aspects for Phase/Track 2/a. Annex A of this TR contains the core part of it to provide the context within which the test results provided below were obtained. The complete document is found for reference in the electronic attachment of this TR.

The purpose of the 4 selection test experiments (Experiments BS1534-1 – BS1534-4) was to evaluate the performance of the IVAS specific ISAR solution candidate with respect to the performance requirements and objectives defined in ISAR TR 26.865 [2].

Table 6.1-1 shows a high-level overview of the experiments. Each experiment was carried out twice (in experiments a and b), once by the solution proponent and once by a cross-checker (XC).

Table 6.1-1: High-level overview of ISAR selection experiments

Exp	Input format	Source material	Listening environment	Bitrates kbps	Listening Lab
BS1534-1a BS1534-1b	SBA (HOA3)	Generic Audio	Headphones	IVAS: 512, CuT: 768	Dolby Qualcomm (XC)
BS1534-2a BS1534-2b	Multi-channel 7.1+4	Generic Audio	Headphones	IVAS: 512, CuT: 768	Fraunhofer Ittiam (XC)
BS1534-3a BS1534-3b	Objects (ISM-4)	Generic Audio	Headphones	IVAS: 512, CuT: 768	Fraunhofer Nokia (XC)
BS1534-4a BS1534-4b	MASA (2 TC)	Generic Audio	Headphones	IVAS: 512, CuT: 768	Dolby Bytedance (XC)

6.1.1.2 Test conditions

A description of the test conditions of all experiments is given in Table 6.1-2.

Table 6.1-2: Description of test conditions

Condition	Description
c01 (REF)	Hidden reference: Native coding system (IVAS@512kbps rendered to post renderer pose)
c02 (LP7)	LP7 anchor: Hidden reference, 7Khz LP filtered
c03 (0DOF)	0-DOF native transcoding reference (IVAS@512kbps binaurally rendered to pre-renderer pose, IVAS stereo coded@256kbps)
c04 (CuT)	3-DOF system under test (IVAS@512kbps split-rendered with ISAR operating at 512kbps)

6.1.1.3 Requirements and Objectives

All experiments check the same requirements defined in TR 26.865 [2], namely that the QoE of the ISAR split rendering system (c04) is no worse than the 0-DOF native transcoding reference system (c03) using the same operation point of the native coding system (IVAS coding at 512 kbps) and best possible operation point for transcoding (IVAS stereo at 256 kbps). The 4 experiments evaluate the requirement for the 4 different main head-trackable IVAS coding formats, i.e., SBA (HOA3), MC 7.1.4, ISM-4 and MASA.

The objectives defined in TS 26.865 [2] is that QoE provided by split rendering solution should be as close as possible to quality of native coding reference system using same operation point. There is no statistical test to verify if this objective is met. However, a statement will be made based on the observed test scores how close the quality of the tested ISAR split rendering solution for the given immersive audio input format is to the quality of the native coding reference system.

6.1.1.4 Test results

6.1.1.4.1 BS-1534-1: SBA (HOA3)

6.1.1.4.1.1 Result plots

Provided below are the result plots for the two BS1534-1 experiments.

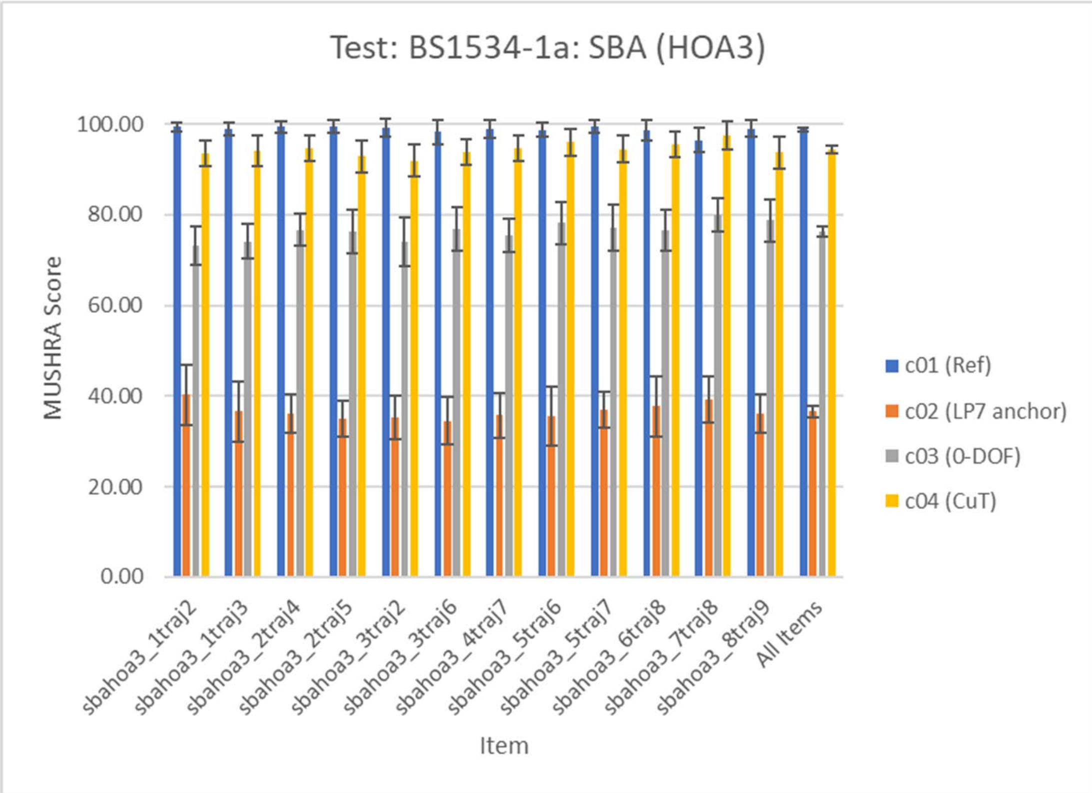

Figure 6.1-1: Results of BS1534-1a test for SBA input audio

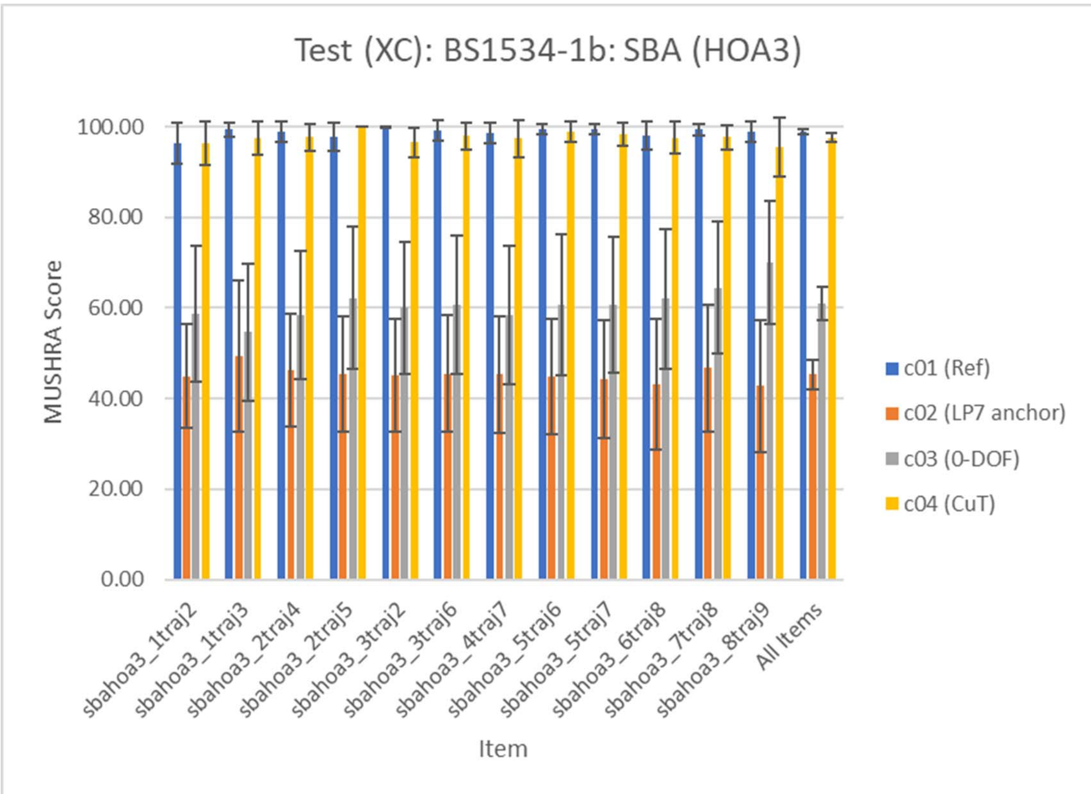

Figure 6.1-2: Results of BS1534-1b test for SBA input audio

6.1.1.4.1.2 Statistical analysis

Provided below is the statistical analysis result for the two BS1534-1 experiments.

Table 6.1-3: Result of statistical analysis of BS1534-1a test checking CuT NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	T	Prob.	ToR
-17.9917	5.7941	0.5289	-34.0156	1.0000	Pass

Table 6.1-4: Result of statistical analysis of BS1534-1b test checking CuT NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	T	Prob.	ToR
-36.6583	20.3336	1.8562	-19.7492	1.0000	Pass

6.1.1.4.1.3 Experimental conclusions

Conclusion of both experiments is that the ISAR split rendering solution for SBA input meets the requirement to be no worse than the 0-DOF transcoding reference system. The experiments indicate that the achievable quality is even clearly better whereby a quality level in the 'excellent' range is achieved compared to the 0-DOF transcoding reference which is providing quality in the 'good' range. The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high 'excellent' range which indicates only very minor audible differences.

6.1.1.4.2 BS-1534-2: Multi-Channel 7.1.4

6.1.1.4.2.1 Result plots

Provided below are the result plots for the two BS1534-2 experiments.

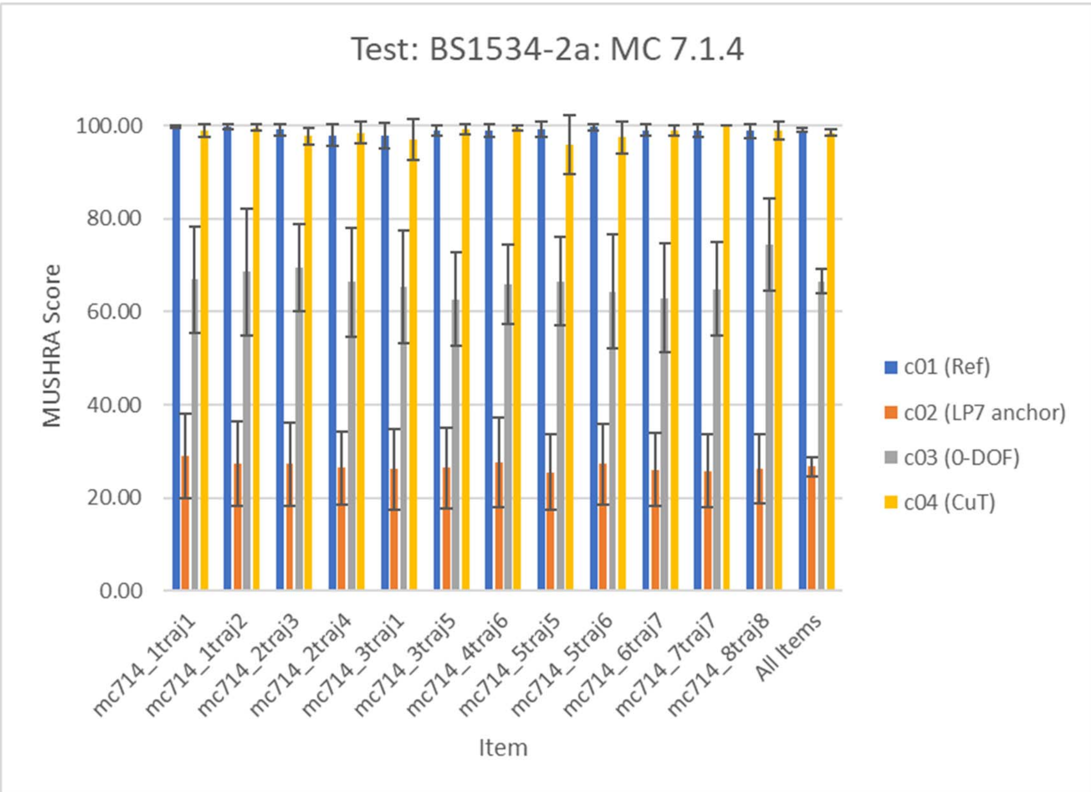

Figure 6.1-3: Results of BS1534-2a test for MC 7.1.4 input audio

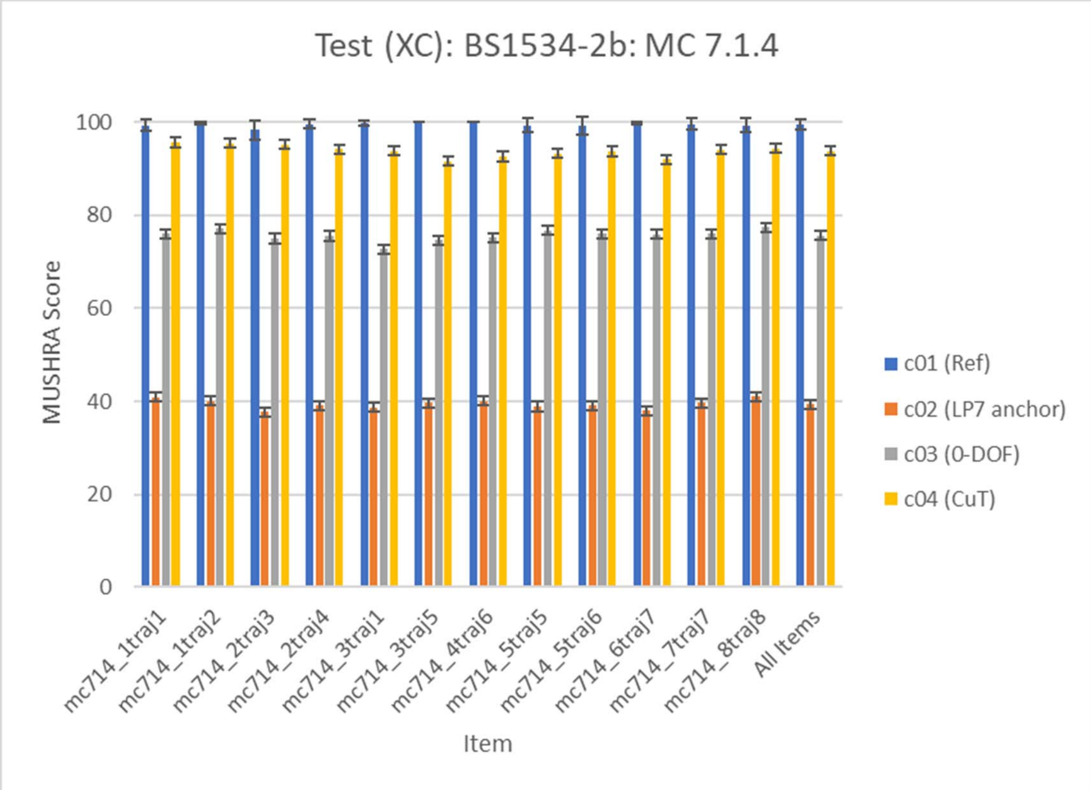

Figure 6.1-4: Results of BS1534-2b test for MC 7.1.4 input audio

6.1.1.4.2.2 Statistical analysis

Provided below is the statistical analysis result for the two BS1534-2 experiments.

Table 6.1-5: Result of statistical analysis of BS1534-2a test checking CuT NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-31.9583	15.1321	1.3814	-23.1353	1.0000	Pass

Table 6.1-6: Result of statistical analysis of BS1534-2b test checking CuT NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-18.1667	6.0881	0.5558	-32.6879	1.0000	Pass

6.1.1.4.2.3 Experimental conclusions

Conclusion of both experiments is that the ISAR split rendering solution for Multi-Channel 7.1.4 input meets the requirement to be no worse than the 0-DOF transcoding reference system. The experiments indicate that the achievable quality is even clearly better whereby a quality level in the ‘excellent’ range is achieved compared to the 0-DOF transcoding reference which is providing quality in the ‘good’ range. The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high ‘excellent’ range, which indicates only very minor audible differences.

6.1.1.4.3 BS-1534-3: ISM-4

6.1.1.4.3.1 Result plots

Provided below are the result plots for the two BS1534-3 experiments.

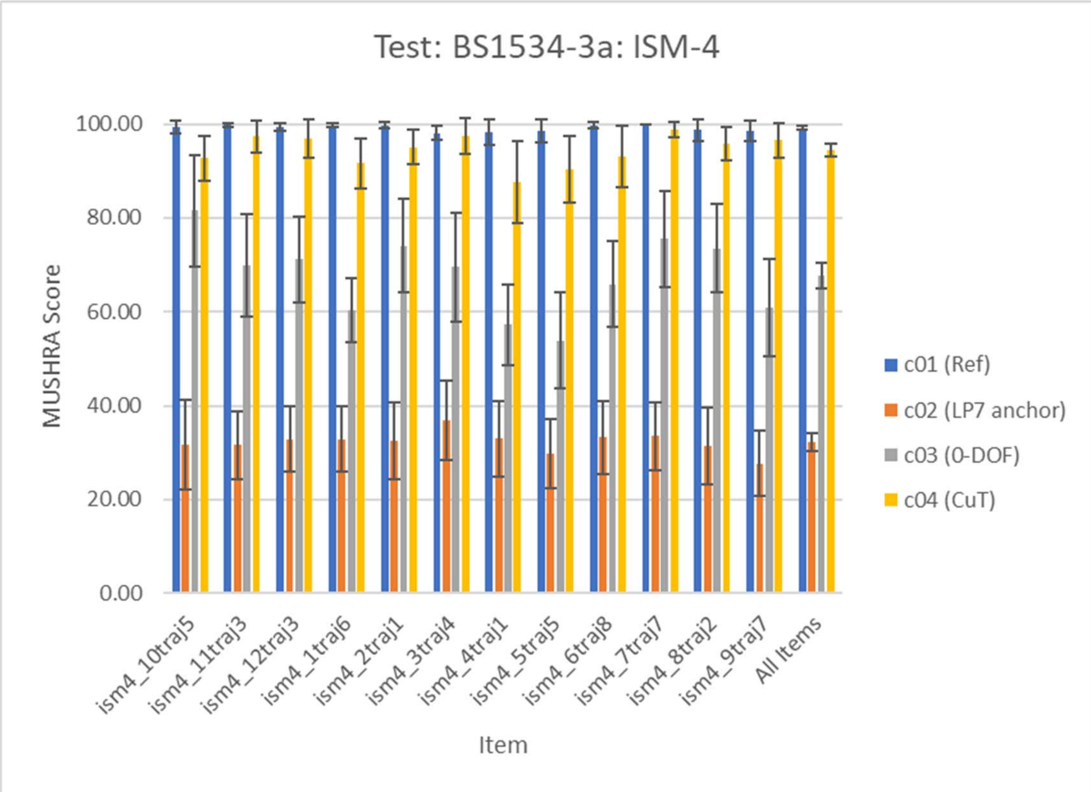

Figure 6.1-5: Results of BS1534-3a test for ISM-4 input audio

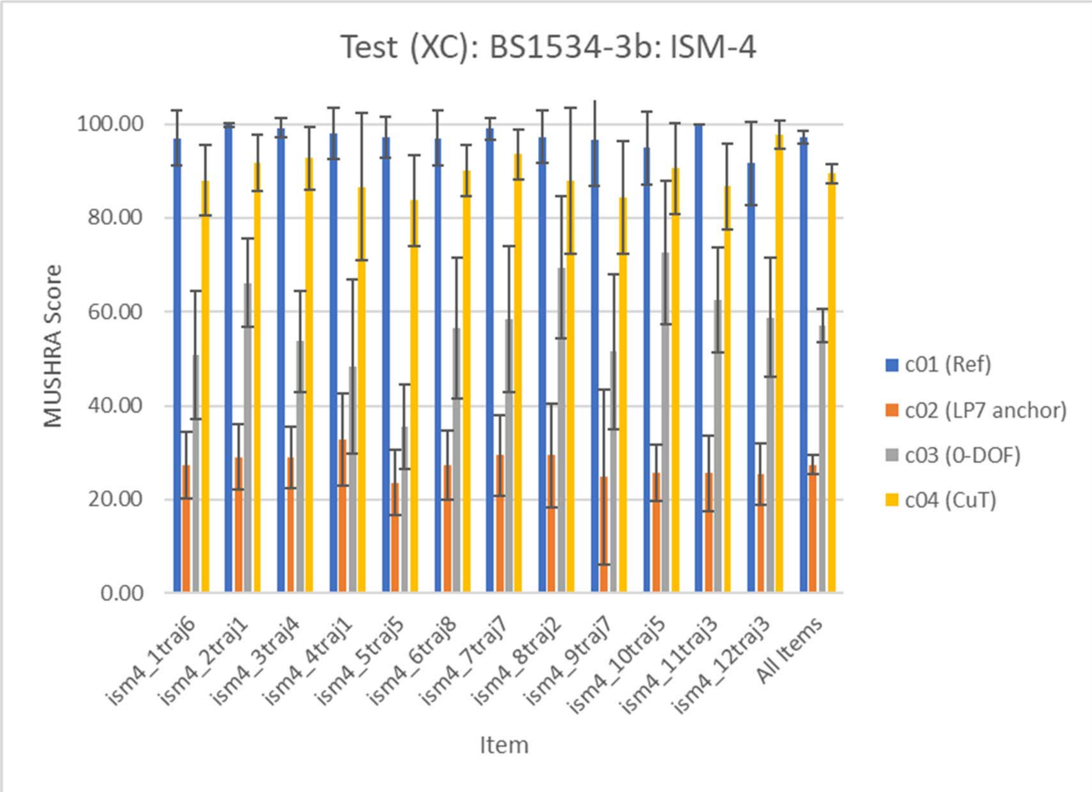

Figure 6.1-6: Results of BS1534-3b test for ISM-4 input audio

6.1.1.4.3.2 Statistical analysis

Provided below is the statistical analysis result for the two BS1534-3 experiments.

Table 6.1-7: Result of statistical analysis of BS1534-3a test checking CuT NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	T	Prob.	ToR
-26.6500	14.7310	1.3448	-19.8178	1.0000	Pass

Table 6.1-8: Result of statistical analysis of BS1534-3b test checking CuT NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-32.4083	22.0761	2.0153	-16.0814	1.0000	Pass

6.1.1.4.3.3 Experimental conclusions

Conclusion of both experiments is that the ISAR split rendering solution for ISM-4 input meets the requirement to be no worse than the 0-DOF transcoding reference system. The experiments indicate that the achievable quality is even clearly better whereby a quality level in the ‘excellent’ range is achieved compared to the 0-DOF transcoding reference which is providing quality in the ‘good’ range. The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high ‘excellent’ range, which indicates only very minor audible differences.

6.1.1.4.4 BS-1534-4: MASA

6.1.1.4.4.1 Result plots

Provided below are the result plots for the two BS1534-4 experiments.

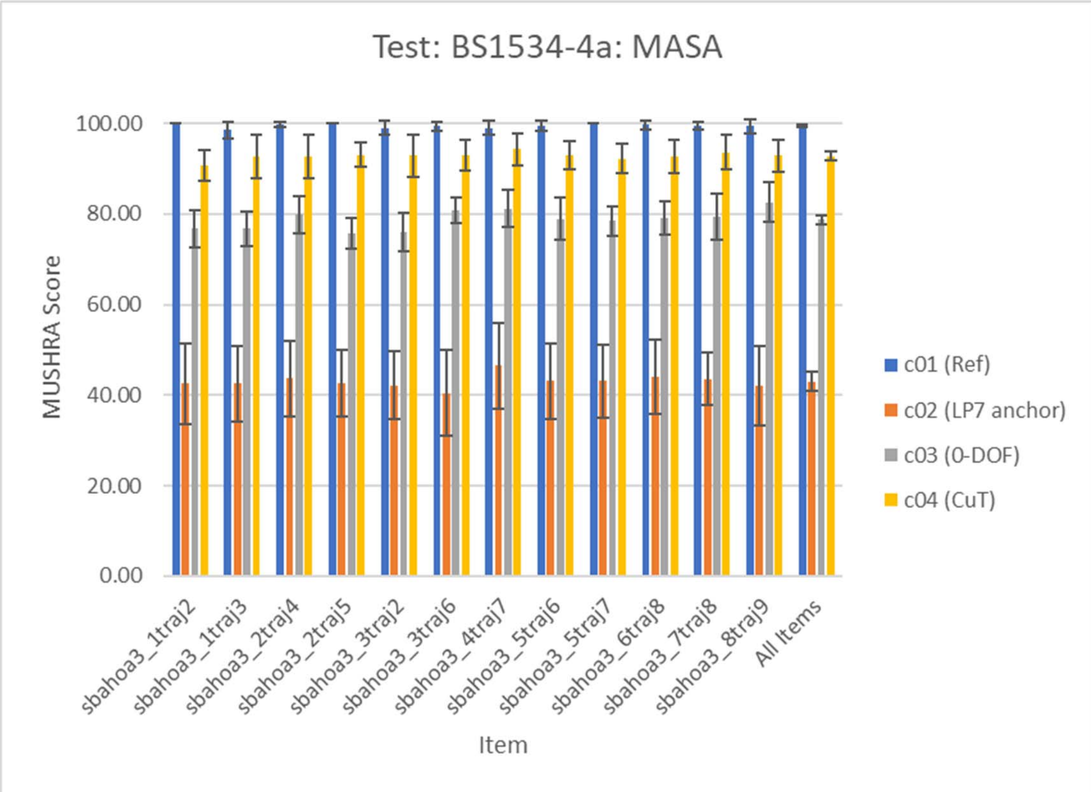

Figure 6.1-7: Results of BS1534-4a test for MASA input audio

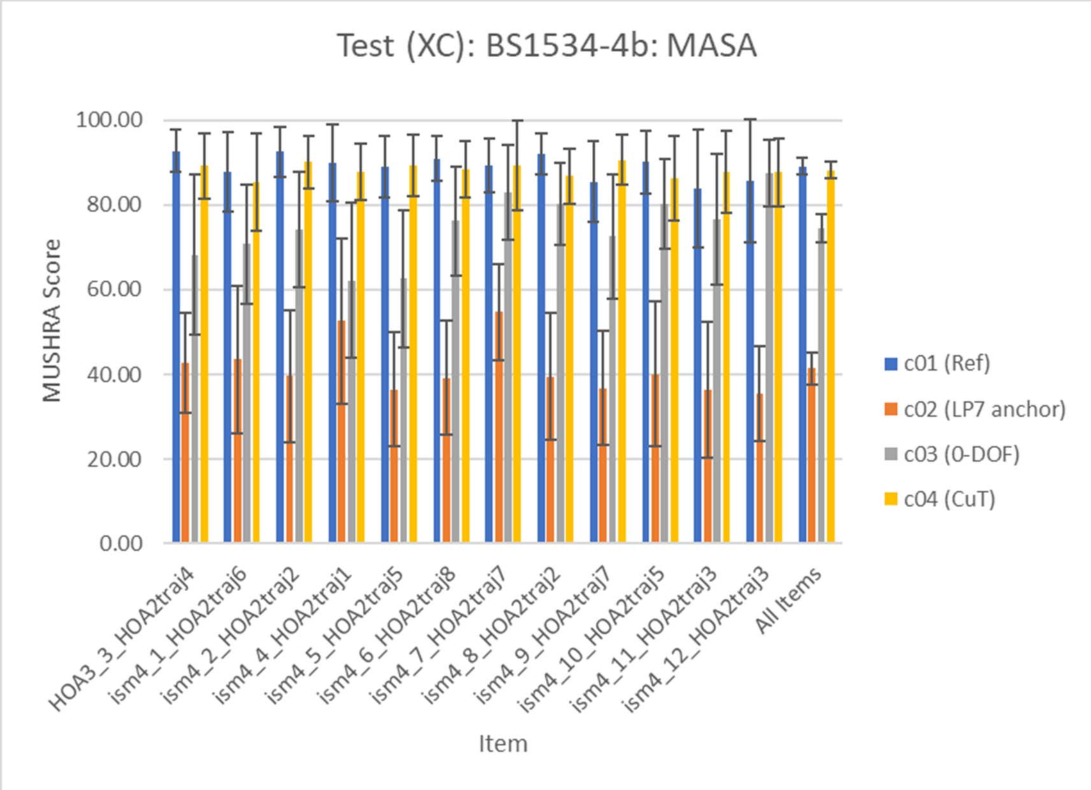

Figure 6.1-8: Results of BS1534-4b test for MASA input audio

6.1.1.4.4.2 Statistical analysis

Provided below is the statistical analysis result for the two BS1534-4 experiments.

Table 6.1-9: Result of statistical analysis of BS1534-4a test checking CuT NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-14.0167	5.0493	0.4609	-30.4091	1.0000	Pass

Table 6.1-10: Result of statistical analysis of BS1534-4b test checking CuT NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	T	Prob.	ToR
-13.7000	22.1500	2.0220	-6.7754	1.0000	Pass

6.1.1.4.4.3 Experimental conclusions

Conclusion of both experiments is that the ISAR split rendering solution for MASA input meets the requirement to be no worse than the 0-DOF transcoding reference system. The experiments indicate that the achievable quality is even clearly better whereby a quality level in the ‘excellent’ range is achieved compared to the 0-DOF transcoding reference which is providing quality in the ‘good’ range. The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high ‘excellent’ range, which indicates only very minor audible differences.

6.1.1.5 Overall conclusion

Conclusion of all 8 experiments testing the requirement that the ISAR split rendering solution for IVAS shall be no worse than the 0-DOF transcoding reference system is that this requirement is met across all tested immersive input audio formats. It can generally be observed that the achievable quality is even clearly better whereby a quality level in the 'excellent' range close to the quality of the native IVAS coding reference system is achieved. In contrast, the 0-DOF transcoding alternative offers substantially lower quality.

The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high 'excellent' range, which indicates only very minor audible differences.

6.1.2 Further tests

6.1.2.1 ISAR tests at 384 and 512 kbps

6.1.2.1.1 Overview

In addition to the selection tests evaluating the performance of the single ISAR candidate solution for IVAS described above, analogous experiments were carried out with the selected ISAR solution operated at 384 and 512 kbps. These experiments were provided by a company contribution and described below. The original contribution is found for reference in the electronic attachment of this TR.

6.1.2.1.2 Test plan

The test plan for the listening experiments of the ISAR solution for IVAS, operated at 384 and 512 kbps (here referred to as low-rate (LR)) is identical with the test plan of the selection tests described in clause 6.1.1.1 with the exception that conditions with the two lower ISAR bit rates are added instead of the 768 kbps condition. The experiments were in only a single instance, i.e., without duplication by a second lab.

The experiments and corresponding input formats are listed in Table 6.1-11.

Table 6.1-11: High-level overview of ISAR low-rate experiments

Exp	Input format	Source material	Listening environment	Bitrates kbps	Listening Lab
BS1534-1-LR	SBA (HOA3)	Generic Audio	Headphones	IVAS: 512, CuT: 384 IVAS: 512, CuT: 512	Dolby
BS1534-2-LR	Multi-channel 7.1+4	Generic Audio	Headphones	IVAS: 512, CuT: 384 IVAS: 512, CuT: 512	Dolby
BS1534-3-LR	Objects (ISM-4)	Generic Audio	Headphones	IVAS: 512, CuT: 384 IVAS: 512, CuT: 512	Dolby
BS1534-4-LR	MASA (2 TC)	Generic Audio	Headphones	IVAS: 512, CuT: 384 IVAS: 512, CuT: 512	Dolby

6.1.2.1.3 Test conditions

A description of the test conditions of all experiments is given in Table 6.1-12.

Table 6.1-12: Description of test conditions

Condition	Description
c01 (REF)	Hidden reference: Native coding system (IVAS@512kbps rendered to post renderer pose)
c02 (LP7)	LP7 anchor: Hidden reference, 7Khz LP filtered
c03 (0DOF)	0-DOF native transcoding reference (IVAS@512kbps binaurally rendered to pre-renderer pose, IVAS stereo coded@256kbps)
c04 (CuT1)	3-DOF system 1 under test (IVAS@512kbps split-rendered with ISAR operating at 512kbps)
c05 (CuT2)	3-DOF system 2 under test (IVAS@512kbps split-rendered with ISAR operating at 384kbps)

6.1.2.1.4 Requirements and Objectives

See clause 6.1.1.3.

6.1.2.1.5 Test results

6.1.2.1.5.1 Experiment BS1534-1-LR: SBA (HOA3)

Provided below is the result plot and tables with statistical analysis results for the BS1534-1-LR experiment.

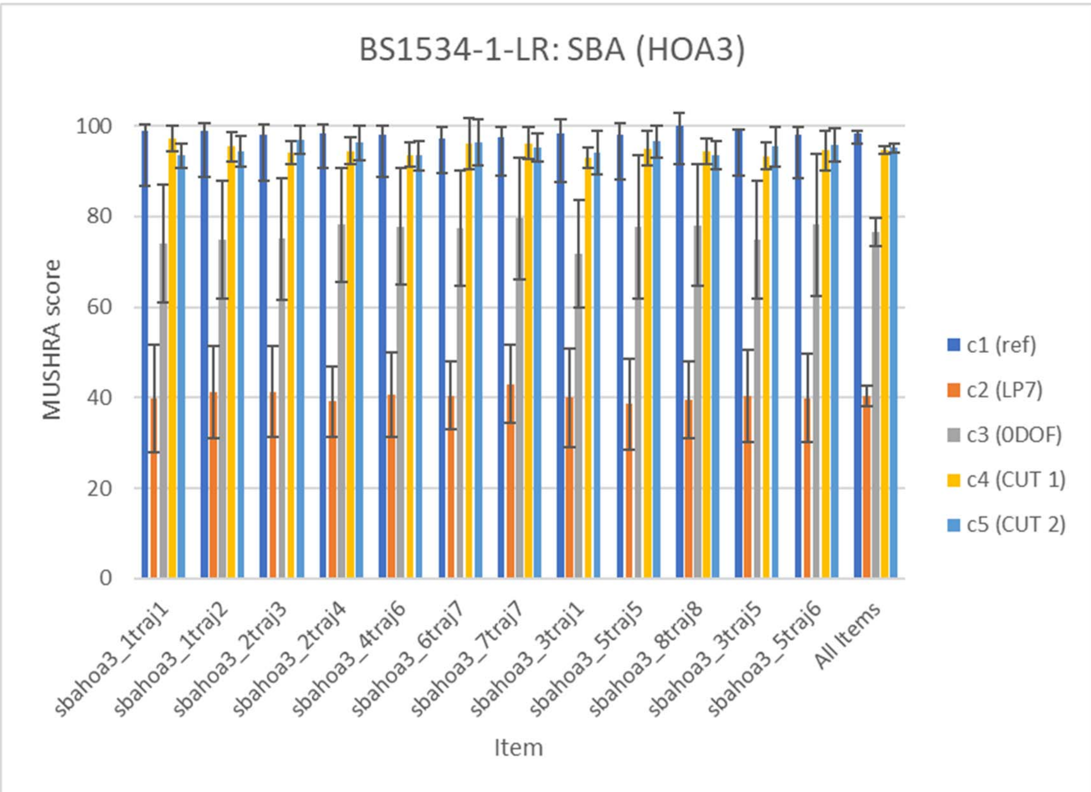

Figure 6.1-9: Results of BS1534-1-LR test for SBA input audio

Table 6.1-13: Result of statistical analysis of BS1534-1-LR test checking CuT1 NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-18.3021	14.2564	1.4550	-12.5784	1.0000	Pass

Table 6.1-14: Result of statistical analysis of BS1534-1-LR test checking CuT2 NWT 0-DOF Reference

Mean Diff. (c03 - c05)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-18.6354	14.3455	1.4641	-12.7280	1.0000	Pass

6.1.2.1.5.2 Experiment BS1534-2-LR: Multi-Channel 7.1.4

Provided below is the result plot and tables with statistical analysis results for the BS1534-2-LR experiment.

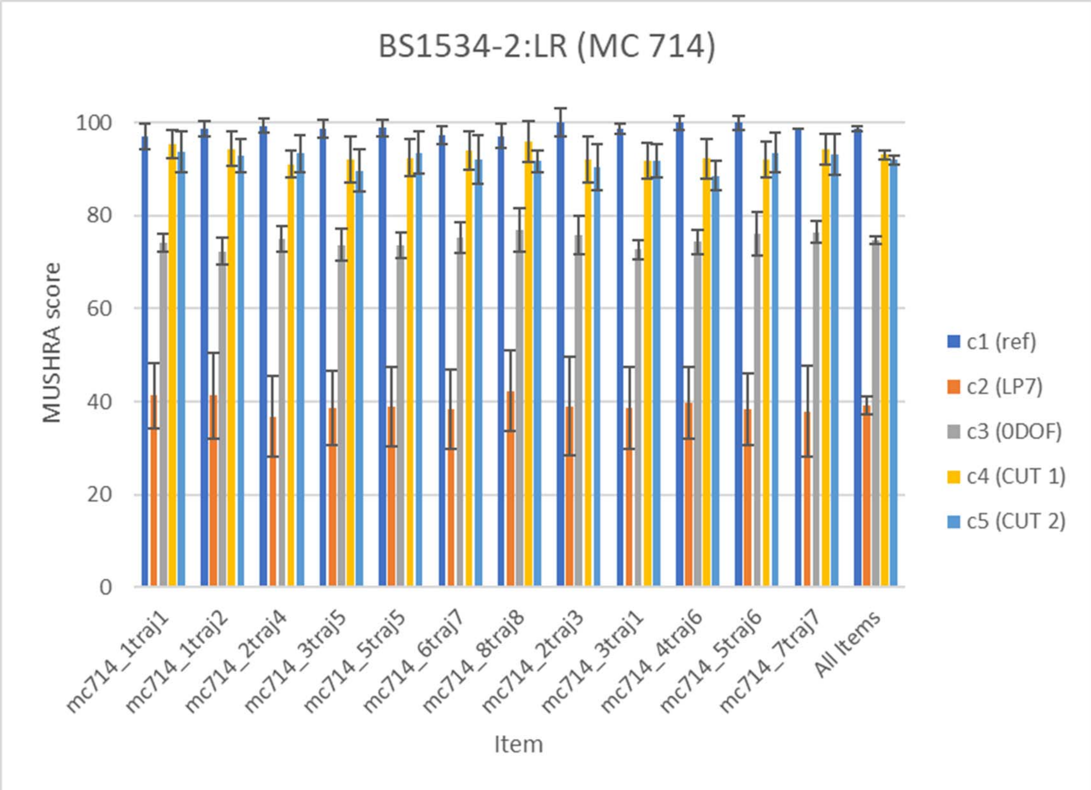

Figure 6.1-10: Results of BS1534-2-LR test for Multi-Channel 7.1.4 input audio

Table 6.1-15: Result of statistical analysis of BS1534-2-LR test checking CuT1 NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-18.4479	4.4506	0.4542	-40.6130	1.0000	Pass

Table 6.1-16: Result of statistical analysis of BS1534-2-LR test checking CuT2 NWT 0-DOF Reference

Mean Diff. (c03 - c05)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-17.3542	5.0304	0.5134	-33.8016	1.0000	Pass

6.1.2.1.5.3 Experiment BS1534-3-LR: ISM-4

Provided below is the result plot and tables with statistical analysis results for the BS1534-3-LR experiment.

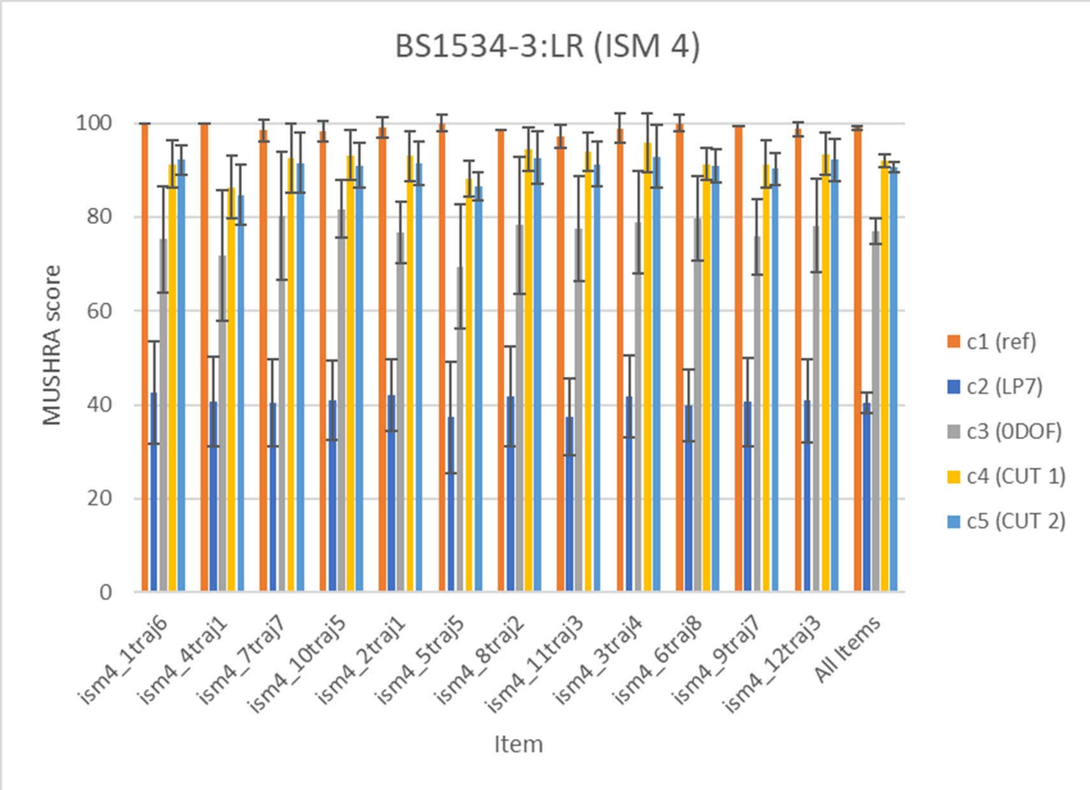

Figure 6.1-11: Results of BS1534-3-LR test for ISM-4 input audio

Table 6.1-17: Result of statistical analysis of BS1534-4-LR test checking CuT1 NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-15.0938	10.4419	1.0657	-14.1629	1.0000	Pass

Table 6.1-18: Result of statistical analysis of BS1534-3-LR test checking CuT2 NWT 0-DOF Reference

Mean Diff. (c03 - c05)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-13.6771	9.9202	1.0125	-13.5085	1.0000	Pass

6.1.2.1.5.4 Experiment BS1534-4-LR: MASA

Provided below is the result plot and tables with statistical analysis results for the BS1534-4-LR experiment.

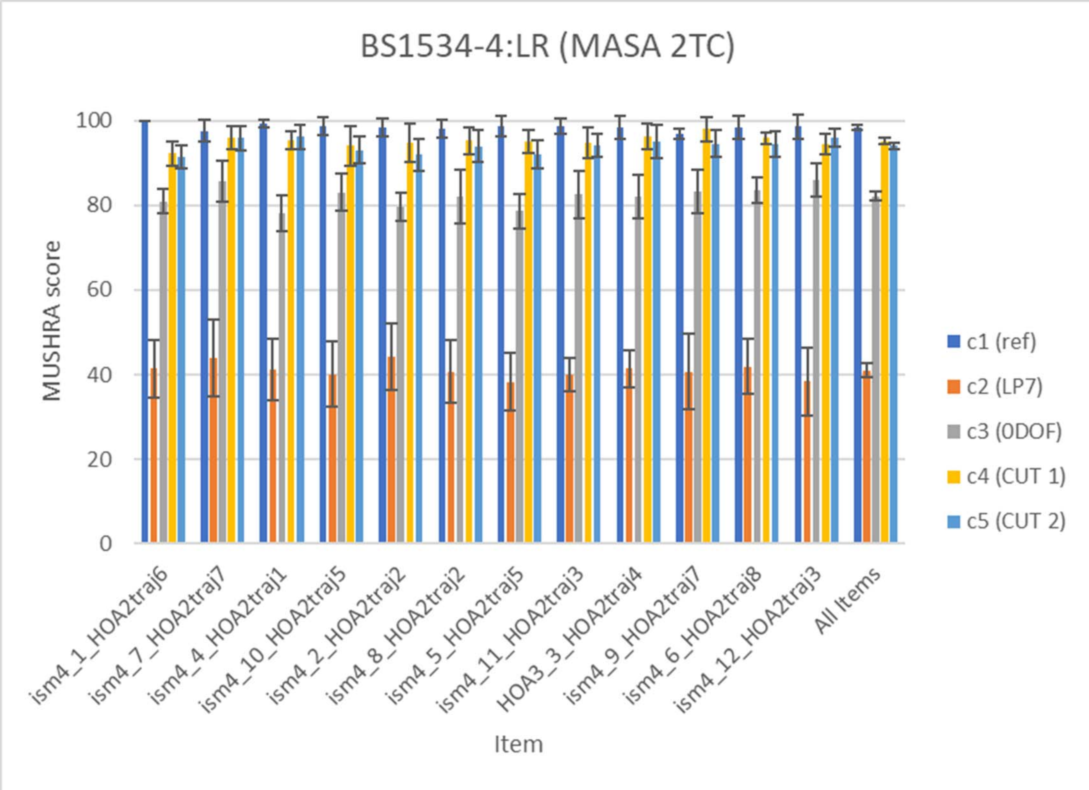

Figure 6.1-12: Results of BS1534-4-LR test for MASA input audio

Table 6.1-19: Result of statistical analysis of BS1534-4-LR test checking CuT1 NWT 0-DOF Reference

Mean Diff. (c03 - c04)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-13.0104	5.7647	0.5884	-22.1132	1.0000	Pass

Table 6.1-20: Result of statistical analysis of BS1534-4-LR test checking CuT2 NWT 0-DOF Reference

Mean Diff. (c03 - c05)	Stdev Diff.	SE_{MD}	t	Prob.	ToR
-11.8958	5.6819	0.5799	-20.5134	1.0000	Pass

6.1.2.2 Overall conclusion

Conclusion of the low-rate experiments is that the ISAR solution for IVAS, when operated at 384 and 512 kbps, respectively, still clearly meets the requirement that it shall be no worse than the 0-DOF transcoding reference system. This is true across all tested immersive input audio formats, SBA, Multi-Channel 7.1.4, ISM-4 and MASA. It can generally be observed that the achievable quality is even clearly better whereby a quality level in the 'excellent' range close to the quality of the native IVAS coding reference system is achieved. In contrast, the 0-DOF transcoding alternative offers substantially lower quality.

The objective to provide a quality level as close as possible to the native coding reference system is met in the sense that the quality score of the split rendering system is in the high 'excellent' range, which indicates only very minor audible differences.

Despite the lower bit rates, this conclusion is the same as for the highest tested bit rate of 784 kbps.

7 ISAR solution properties

7.1 ISAR Baseline properties

7.1.1 Introduction

The properties of the ISAR baseline are best characterized by comparing the reported properties of the selected split rendering solution against the design constraints. The design constraints are specified in TR 26.865 [2] clauses 6 (Physical Design Constraints) and 7 (Functional Design Constraints).

To that end, in the following, relevant excerpts of the ISAR selection deliverable on compliance with design constraints are provided. The full deliverables document is provided in the electronic attachment of this TR for reference, e.g., to help understanding how the solution properties were determined.

7.1.2 Physical properties of ISAR baseline in comparison to design constraints

The following table 7.1-1 displays the physical properties of the ISAR baseline in relation to the physical design constraints:

Table 7.1-1: Physical properties of the ISAR baseline in relation to physical design constraints

Physical attribute	Constraint	Measured values based on "isar_selection_branch"(1
Complexity of operation in end-rendering lightweight device	The complexity of operation in end-rendering lightweight device shall not exceed 80+(DOF*20) wMOPS. 0DOF: 80 1DOF: 100 2DOF: 120 3DOF: 140	3DOF (with LC3plus): max 73.66, avg 70.7 3DOF (with LCLD): max 60.96, avg 51.41
Complexity of operation at capable device/node	The complexity of operation at pre-rendering device/node shall be characterized.	IVAS decoder + 3DOF split pre-rendering with LC3plus (MC714): max 1063.97, avg 1037.15 IVAS decoder + 3DOF split pre-rendering with LC3plus (ISM4): max 864.88, avg 854.2 IVAS decoder + 3DOF split pre-rendering with LCLD (HOA3): max 925.5, avg 837.0 IVAS decoder + 3DOF split pre-rendering with LCLD (MASA2): max 662.12, avg 653.06
Memory footprint of operation in end-rendering lightweight device	The RAM consumption shall not exceed 100+(DOF*50) kWords. 0DOF: 100 kWords 1DOF: 150 kWords 2DOF: 200 kWords 3DOF: 250 kWords The ROM (PROM and table ROM) shall not exceed 150 kWords.	(word = 4 Bytes) RAM (heap + stack): 3DOF (with LC3plus): 67.51 kWords 3DOF (with LCLD): 67.87 kWords PROM (lc3plus + lib_isar + lcld): 20.04 kWords TROM (lc3plus + lib_isar + lcld): 42.14 kWords
Memory footprint of operation at pre-rendering device/node	The memory footprint of operation at pre-rendering device/node shall be characterized.	RAM: IVAS decoder + 3DOF split pre-rendering with LC3plus (MC714): 586.751 kWords IVAS decoder + 3DOF split pre-rendering with LC3plus (ISM4): 393.24 kWords IVAS decoder + 3DOF split pre-rendering with LCLD (HOA3): 467.63 kWords IVAS decoder + 3DOF split pre-rendering with LCLD (MASA2): 290.39 kWords PROM (lib_com + lib_dec + lib_rend):172.6 kWords TROM (lib_com + lib_dec + lib_rend):156.4 kWords
Algorithmic motion-to-sound latency in head-tracked rendering operation	0-DOF: no constraint 1-DOF: 30 ms for rotations around corrected axis (in post rendering), for other axes no constraints 2-DOF: 30 ms for rotations around corrected axes (in post rendering), for the remaining axis no constraint 3-DOF: 30 ms for rotations around all axes (in post rendering)	2.5ms <= Algorithmic motion-to-sound latency <= 22.5 ms

<table>
<tr>
<td>Algorithmic audio delay</td>
<td>The total algorithmic end-to-end audio delay including IVAS algorithmic delay shall not exceed 50 ms.</td>
<td>
<table>
<tr><th></th><th>LC3plus</th><th>LCLD</th><th>DOF</th></tr>
<tr><td>SBA@512kbps</td><td>38+5+2.5 = 45.5ms</td><td>38ms</td><td>3</td></tr>
<tr><td>MC714@512kbps</td><td>32+5+2.5 = 39.5ms</td><td>32+5 = 37ms</td><td>3</td></tr>
<tr><td>ISM@512kbps</td><td>32+5+2.5 = 39.5ms</td><td>32+5 = 37ms</td><td>3</td></tr>
<tr><td>MASA@512kbps</td><td>38+5+2.5 = 45.5ms</td><td>38ms</td><td>3</td></tr>
<tr><td>OMASA@512kbps</td><td>38+5+2.5 = 45.5ms</td><td>38ms</td><td>3</td></tr>
<tr><td>OSBA@512kbps</td><td>38+5+2.5 = 45.5ms</td><td>38ms</td><td>3</td></tr>
</table>
<table>
<tr><th></th><th>LC3plus</th><th>LCLD</th><th>DOF</th></tr>
<tr><td>SBA@512kbps</td><td>38+2.5 = 40.5ms</td><td>38ms</td><td>0</td></tr>
<tr><td>MC714@512kbps</td><td>32+2.5 = 34.5ms</td><td>32+5 = 37ms</td><td>0</td></tr>
<tr><td>ISM@512kbps</td><td>32+2.5 = 34.5ms</td><td>32+5 = 37ms</td><td>0</td></tr>
<tr><td>MASA@512kbps</td><td>38+2.5 = 40.5ms</td><td>38ms</td><td>0</td></tr>
<tr><td>OMASA@512kbps</td><td>38+2.5 = 40.5ms</td><td>38ms</td><td>0</td></tr>
<tr><td>OSBA@512kbps</td><td>38+2.5 = 40.5ms</td><td>38ms</td><td>0</td></tr>
</table>
NOTE: The additional delay for 1-3DOF on top of IVAS is 5ms for pose correction with LCLD coding, 7.5ms (5+2.5ms) for pose correction and LC3plus coding. The additional delay on top of IVAS for 0-DOF operation is 2.5ms for LC3plus coding and 0 or, respectively, 5ms for LCLD coding. In case of pre-rendering in CLDFB domain additional delay by LCLD coding can be circumvented.
</td>
</tr>
<tr>
<td>Bit rate of coded intermediate representation</td>
<td>The Split Rendering solution should offer 3-DOF operation at multiple bit rates with rate switching support and shall at least offer operation at 768 kbps.
For 3-DOF operation, 384 kbps and 512 kbps should be offered in addition to the required 768 kbps.
For 1-DOF operation, 256 kbps should be offered.
For 0-DOF operation, 256 kbps and even lower bit rates should be offered.
For cases with high transmission channel capacities, high bit rate operation modes > 768 kbps may be offered if QoE gains can be demonstrated.</td>
<td>
<table>
<tr><th>DOF</th><th>Bitrates supported (kbps)</th></tr>
<tr><td>0</td><td>256, 384, 512</td></tr>
<tr><td>1</td><td>384, 512, 768</td></tr>
<tr><td>2</td><td>384, 512, 768</td></tr>
<tr><td>3</td><td>384, 512, 768</td></tr>
</table>
</td>
</tr>
</table>

7.1.4 Supported functional features in relation to functional design constraints

The following table 7.1-2 lists supported functional features of the ISAR baseline in relation to the functional design constraints.

Table 7.1-2: Supported functional features of the ISAR baseline in relation to the functional design constraints

Functional attribute	Constraint	Supported/Unsupported
Immersive audio formats of native coding format	All required immersive IVAS encoder input formats according to Pdoc IVAS-4 shall be supported. IVAS stereo input format should be supported for 0-DOF operation.	Supported: SBA MC ISM MASA OSBA OMASA Unsupported: Stereo
Bit rates of required immersive audio coding modes of native coding format	All required bit rates of IVAS immersive operation modes according to Pdoc IVAS-4 shall be supported. All IVAS stereo bit rates should be supported for 0-DOF operation.	Supported: All bitrates supported for immersive formats Unsupported: Stereo
Head-trackability of immersive audio formats	The head-trackability of the immersive audio formats of the native coding format shall be retained. Preservation of the DOF level is the objective, reduced DOF levels may be provided. Note: an audio format supported by the native coding system may be 3-DOF head-trackable, i.e., around 3 axes (yaw, pitch, roll). The Split Rendering system should retain this possibility. Complexity or bit rate reduced variants may though reduce this to lower DOF levels like yaw-only correction (1-DOF).	Supported
Packet loss concealment (PLC)	A PLC solution shall be provided.	Supported
Non-diegetic audio support	The solution shall support (1DOF – 3DOF) diegetic audio and (0-DOF) one-channel non-diegetic audio and two-channel (stereo or binaural) non-diegetic audio. It shall be possible to overlay post rendered audio obtained from instances operated with diegetic and non-diegetic audio.	Supported

Annex A (informative): ISAR Pdoc on Testing Aspects for Phase/Track 2/a

A.1 Overview

This Annex contains the core part of the permanent document on Testing Aspects for the ISAR Phase/Track 2/a solution selection. The complete document is found for reference in the electronic attachment of this TR.

A.2 Organization of tests

A.2.1 Overview of the Selection Test Process

The selection tests of the IVAS specific ISAR solution will be organized as in-house tests by Solution Proponents Laboratories. The execution of these subjective tests is under the responsibility of the solution proponents participating in the selection and other volunteering organizations.

The selection test experiments will be duplicated and additionally run by suitable cross-checkers, Cross-check Listening Laboratories, with no stake in the candidate solutions under test.

The selection test results will be reported by the testing organizations to SA4 with suggested statistical result analysis. SA4 will review these analyses and, if found valid, confirm them to base its selection decision on them.

The processing of the selection test material is under the responsibility of the solution proponents participating in the selection. It is based on commonly available processing scripts, solution candidate executables, original sound material and head-tracker trajectories available to the solution proponents and other organizations who volunteer to carry out cross-checks.

A.2.2 Responsibilities

A.2.2.1 Solution Proponent (SP)

The specific responsibilities of the SP are:

- Make executables of solution candidate publicly available.
- Develop common processing scripts using the condition lists defined in this document and the processing steps defined in the processing plan.
- Process the test material using commonly available processing scripts, the shared solution candidate executables, original sound material and head-tracker trajectories.
- Communicate with Volunteering Processing Cross-check Organizations to verify correct processing.

A.2.2.2 Solution Proponents Laboratory (SPL)

- Carry out selection tests according to the requirements of this test plan.
- Carry out statistical result analysis according to the requirements of this test plan and provide test report including analysis to SA4.
- Obligations as SPL:
    - The testing shall be caried out in a blinded fashion not revealing the conditions to the subjects.
    - No test subjects must be used that were actively involved in developing the split rendering features in the systems under test that are exposed by the experiments.

  - The test report shall describe how the listening lab ensured unbiased testing.

A.2.2.3 Volunteering Cross-check Listening Laboratories (CLL)

- Carry out selection tests according to the requirements of this test plan.
- Carry out statistical result analysis according to the requirements of this test plan and provide test report including analysis to SA4.
- Obligations as CLL:
  - The testing shall be caried out in a blinded fashion not revealing the conditions to the subjects.
  - The CLL shall not be contributor of the split rendering features in the systems under test that are exposed by the experiments.
  - The test report shall contain a statement confirming that the listening lab has met the obligations.

A.2.2.4 Listening Laboratories (LL) (both SPL and CLL)

- Provide a listening environment meeting the listening conditions for BS.1534 testing [12].

A.2.2.5 Volunteering Processing Cross-check Organizations (PCO)

- Process the test material using commonly available processing scripts, the shared solution candidate executables, original sound material and head-tracker trajectories.
- Communicate with Solution Proponent to verify correct processing.

A.2.2.6 SA4

- Review selection test analyses received from LLs and determine their validity.
- Selection of Candidate Solution according to selection rules for IVAS specific ISAR solutions targeted in Phase/Track 2/a of the ISAR Work Plan.

A.2.3 Statistical analysis of results

The statistical result analysis reports shall present the results of the Terms of Reference (ToR) tests using Student's Dependent Groups t-test (single-sided at 95% confidence level). Results of the Requirement ToR tests for each experiment shall be presented containing all relevant data allowing to verify proper execution of the Student's Dependent Groups t-test.

In the for **Requirement ToR tests** this should lead to the following indications:

- Requirement ToR tests that are passed, (i.e., CuT "not worse than" Requirement) are indicated by **CuT NWT Ref**.
- Requirement ToR tests that are exceeded, (i.e., CuT "better than" Requirement) are indicated by **CuT BT Ref**.
- Requirement ToR tests that are failed (i.e., CuT "worse than" Requirement) are indicated by **CuT WT Ref**.

A.3 Identities

A.3.1 Involved organizations

In the following, the identities of the involved organizations are listed:

- SP (proponent of CuT):
  - Proponent companies:

    Dolby Sweden AB, Ericsson LM, Fraunhofer IIS, Nokia Corporation, NTT, Orange, Panasonic Holdings Corporation, Philips International B.V., Qualcomm Incorporated, VoiceAge Corporation
  - Main contributor(s) to CuT whose split rendering features are exposed in the experiments:

    Dolby Sweden AB, Fraunhofer IIS
- SPLs:
  - Dolby Sweden AB, Fraunhofer IIS
- CLLs:
  - Qualcomm, Nokia, Bytedance, Ittiam
- Processing Cross-check Organizations (PCO)
  - Fraunhofer IIS, Dolby

A.3.2 LL assignment

Table A.1-1: Assignment of experiments to listening labs

Experiment	SPL	CCL/other SPL
BS1534-1: SBA (HOA3)	Dolby	Qualcomm
BS1534-2: Multi-channel 7.1+4	Fraunhofer IIS	Ittiam
BS1534-3: Objects	Fraunhofer IIS	Nokia
BS1534-4: MASA	Dolby	Bytedance

A.4 Information relevant to all Experiments

A.4.1 General Technical Notes

Any and all deviations from the specifications contained in this document must be documented and submitted to SA4 along with the test reports.

A.4.2 General Consideration of Experiments

A.4.2.1 Difference scenario between assumed and actual end-device poses

For the evaluation of ISAR split rendering solution, a primary focus should be the testing with relevant difference scenarios between assumed and actual end-device poses. To cover relevant cases, the used head-tracker trajectory files should be taken from the following categories:

- Static within range: +-20 degrees
- Dynamic within range: +-20 degrees
  - Sinusoidal: 0.25 Hz
  - Triangular: 0.5 Hz
- Real, i.e., derived from real head tracker trajectories with movements giving rise to substantial differences (>15 degrees) between assumed and actual end-device poses and exposing the tested methods to a sufficient degree.

A.4.2.2 DOF

Another aspect of interest is the ability to deal with differences of assumed and actual end-device poses around different axes, i.e., the number of degrees of freedom (1-3) which the candidate solutions can cope with. Accordingly, the head-tracker trajectory files shall cover the following DOF scenarios:

- 1-DOF with pose deviations in yaw
- 2-DOF with pose deviations in yaw and pitch
- 3-DOF with pose deviations in yaw, pitch and roll

A.4.2.3 Rendering simulation

Two different rendering simulation methodologies shall be covered in the tests.

- Trajectory nullification

  This simulation methodology is based on the concept that an immersive audio scene is pre-rotated prior to IVAS encoding while the head-tracked rendering compensates for the pre-rotation. In the ideal case and under certain conditions, this compensation can be perfect. This simulation methodology exposes the ability of a split rendering system to compensate for the pre-rotation despite the differences between assumed and actual end-device poses.

- Unguided end-device pose

  This simulation methodology is based on rendering a decoded an immersive audio scene according to a given head-tracker trajectory. This simulation methodology exposes the ability of a split rendering system to follow the actual head-tracker trajectory despite the mere availability of the divergent assumed head-tracker trajectory available at the pre-renderer.

A.4.2.4 Input formats

According to the ISAR requirements, the tests shall cover the following IVAS codec input formats:

- SBA (HOA3)
- MASA (2 TCs)
- Multi-channel (7.1.4)
- Objects (ISM-4)

A.4.3 Methodology

BS.1534 test methodologies shall be used in the ISAR selection tests. High-level configuration of the experiments is outlined below.

- Number of items per experiment: 12
- 10 experienced listeners
- Total number of conditions: 4
- Number of anchor conditions: 2
  - Native reference system
  - 7 kHz low-pass anchor

A.4.4 Head-tracker trajectories

A.4.4.1 General

All head-tracker trajectory files shall follow the convention imposed by the IVAS source code specification [11], i.e., shall be usable by the IVAS decoder/renderer.

A.4.4.2 Head-tracker trajectory categories and DOF

Head-tracker trajectories shall meet the above-defined category and DOF specifications.

A.4.4.3 Head-tracker trajectory availability and selection

Head-tracker trajectories of the above-defined categories and DOF will be publicly collected from volunteering organizations. After collection and checking suitability, a list of available trajectories will be generated. The collection and checking will be done jointly by the involved organizations of the ISAR selection, i.e., the SP, the LLs and the PCOs.

In a second step, these organizations will jointly select up to 6 suitable trajectories for each test. In case this number of trajectories is not available, a smaller number is selected where a given trajectory may be reused across different tests or within a test.

The involved organizations will document their trajectory selection and assignment to tests for inclusion of this information into this document.

A.4.5 Audio Material

A.4.5.1 General

All audio material shall be sampled at 48 kHz with Full Band (FB) content and formatted as 16-bit little endian WAVE format files.

A.4.5.2 Audio categories

To cover a broad range of conceivable audio categories, the test items should be taken from the categories clean and noisy speech, music, critical audio. However, the ability of the system to deal with different audio categories is only a secondary focus of the selections tests and full coverage of these categories may not be possible.

A.4.5.3 Test Item availability and selection

Audio material of the above categories has been collected as part of the IVAS codec selection phase. Details of this material are available in the IVAS test plan (see TR 26.997 [13], Annex E). A subset of this material is either publicly available or at least available to the involved organizations of the ISAR selection, i.e., the SP, the LLs and the PCOs. These organizations will in a first step create a list of commonly available test items.

In a second step, these organizations will jointly select up to 12 suitable original test items for each test. In case this number of test items is not available, a smaller number is selected where a given test item may be reused across different tests or even within a test if the applied head-tracker trajectories, DOF or simulation methodology is different.

The involved organizations will document their test item selection and assignment to tests for inclusion of this information into this document.

A.4.5.4 Training material

No dedicated training material will be made available for use in a potential training phase in which the subjects may familiarize with the testing methodology and environment.

Such a training phase is voluntary and upon the own responsibility of the involved LLs. No items from the main tests shall be used for training. A training phase shall be executed as a separate short BS.1534 session.

A.4.6 Listening Systems and Listening Environments

The ISAR Selection Test will use the following listening systems:

- High-quality stereo headphones for binaural listening, e.g.:
  - Sennheiser HD 650

Annex B: Change history

Change history							
Date	**Meeting**	**TDoc**	**CR**	**Rev**	**Cat**	**Subject/Comment**	**New version**
2024-05	SA4#128	S4-241071				Initial version	0.0.1
2024-05	SA4#128	S4-241133				More complete version by rapporteur prior to SA4#128 meeting	0.0.2
2024-05	SA4#128	S4-241193				Editorial enhancements Addition of ISAR low-rate test results Removal of placeholder of track B test results	0.1.0
2024-05	SA4#128	S4-241346				Editorial updates	0.2.0
2024-06						Version 1.0.0 created by MCC	1.0.0
2024-06						Version 18.0.0 created by MCC upon approval	18.0.0
2024-09	SA#105	SP-241114	0001	2	F	Editorial corrections	18.1.0

History

Document history		
V18.0.0	August 2024	Publication
V18.1.0	October 2024	Publication