ETSI TR 126 950 V10.0.0 (2011-04)

Technical Report

Universal Mobile Telecommunications System (UMTS);
LTE;
Study on Surround Sound for PSS and MBMS
(3GPP TR 26.950 version 10.0.0 Release 10)

Reference

DTR/TSGS-0426950va00

Keywords

LTE, UMTS

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 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Copyright Notification

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is 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 (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, 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.

Foreword

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, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

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

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.

1 Scope

The present document investigates the potential user experience benefits of surround audio in 3GPP services. The investigation will be performed as follows:

- Identify and document relevant use cases for surround sound in 3GPP.
- Define design constraints that would need to be met by a surround audio codec extension method for adoption by 3GPP.
- Identify suitable testing methodology for surround sound in relevant use cases of the PSS and MBMS services.
- Define subjective minimum performance criteria that would need to be met in order to motivate the consideration of a surround audio coding extension for adoption by 3GPP.
- Validate the user benefits and the feasibility of the deployment of surround sound for the PSS and MBMS services according to the defined minimum performance criteria, bit rate and design constraints for all the use cases (such as surround sound speaker set-up and headphone decoding mode) through evaluation of at least one example of surround sound coding methods which may be MPS.

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 TS 26.346: "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs".

[3] 3GPP TS 26.234: "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs".

[4] ITU-R Recommendation BS.775-2: "Multichannel stereophonic sound system with and without accompanying picture," Jul. 2006.

[5] ITU- Recommendation BS.1534-1: "Method for the subjective assessment of intermediate quality level of coding systems", Geneva, 2003.

[6] ISO/IEC JTC1/SC29/WG11 N2006 "Report on the MPEG-2 AAC Stereo Verification Tests"; Feb 1998, http://www.chiariglione.org/mpeg/working_documents/mpeg-02/audio/AAC_results.zip.

[7] 3GPP TR 26.936: "Performance characterization of 3GPP audio codecs".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

HRTF: A Head-Related Transfer Function (HRTF) represents a pair of filters that are obtained by measurement or modelling. It represents the acoustic transmission from a point in space to the entrance of a listener's ear canal. It depends on the relative positions of the source and the listener's head.

3.2 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].

5.1ch	Loudspeaker set-up with 2 front channels, 2 rear channels, 1 center channel and 1 subwoofer
HRTF	Head-Related Transfer Function
MPS	MPEG Surround
MUSHRA	MUlti Stimulus test with Hidden Reference and Anchor

4 Use cases

The relevant use cases considered in this study are applications in the context of MBMS and/or PSS services.

In the home entertainment industry the de facto standard for surround sound content is the 5.1 channel format. The reproduction of such surround signal can be done in various ways using a number of channels that is not necessarily equal to the content at the service provider side resulting in different listening modes. The general characteristics of MBMS and PSS services apply and will be considered to derive design constraints and performance requirements for the study item.

We have identified the following use cases for consideration.

Table 1: List of use cases considered in the study

Use case #	Reproduction	Description
1 a	Headphones	Surround decoding with binaural post-processing
1 b	Headphones	Surround decoding with binaural processing being part of the decoding process
2.1 a	Loudspeakers	Surround decoding followed by rendering on the UE
2.1 b	Loudspeakers	Surround decoding with rendering being part of the decoding process on the UE
2.2	Loudspeakers	Surround bit-stream is transported via the UE. Decoding and rendering is performed in a non-3gpp device connected to the UE.
2.3	Loudspeakers	Surround decoding on the UE. Decoded surround audio data are transported to a non-3gpp device connected to the UE for rendering.

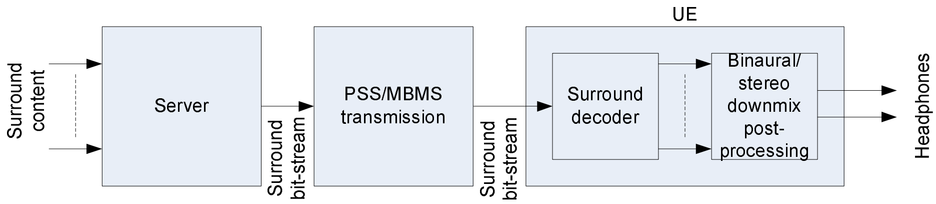
NOTE:

- In the following use cases it is assumed that the surround sound content provided to the server comprises multiple channels, typically 6 channels in the 5.1 format.

- Alternatively, the surround sound content may be presented to the server as a binauralized stereo signal. In this case, the server would encode the surround sound as an artistic downmix (which is also referred to as Binaural Virtual Surround effect). No additional processing would be required when listening over headphones. However, this alternative format would have several implications:
    - When playing over stereo or multichannel loudspeakers, the decoder would have to remove the binauralization effect. Some signalling would be needed to indicate that the downmix is binauralized stereo signal.
    - This alternative format would not offer mono/stereo backward compatibility to existing 3GPP audio codecs, especially when listening over loudspeakers.
- In the following use cases it is assumed that the surround bit-stream contains spatial information to control the behaviour of the surround decoder. The surround decoder produces surround sound based on this side information. However, a possible additional function of the surround capable UE is that the surround decoder may be able to upmix stereo signals encoded by legacy 3GPP audio codecs, which can then be binauralized for listening over headphones.

4.1 Surround sound over headphones

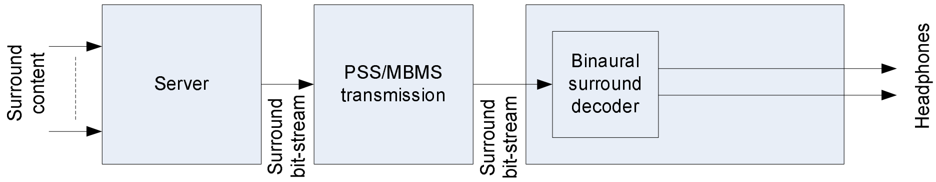
Binaural/Stereo post-processing may or may not be part of the surround sound decoder (see Figures 1 and 2). Figure 1 illustrates a block diagram where the binaural or stereo post-processing is not part of the surround decoder. A server transmits surround sound bit-streams via PSS or MBMS protocols/services. The UE first decodes the received surround bit-stream to a surround signal. The resulting surround signal is processed by binaural or stereo downmix post-processing to produce a stereo signal. The resulting signal can be represented on headphones.



NOTE: The surround bit-stream is decoded inside the UE to a surround signal. This surround signal is input to a binaural or stereo downmix post-processor that produces a representation of the surround signal for headphone reproduction.

Figure 1: Signal flow for use case 1 a where binaural and stereo downmix post-processing is not part of the surround sound decoder

Figure 2 provides a block diagram where binaural post-processing is part of, i.e. integrated into, the surround decoder. The only difference with regard to Figure 1 is that the surround bit-stream is not first decoded to a full surround signal prior to binaural post-processing. Instead the steps of surround decoding and binaural decoding are integrated into a single binaural surround decoder.



NOTE: The surround bit-stream is decoded inside the UE directly to produce a representation of the surround signal for headphone reproduction.

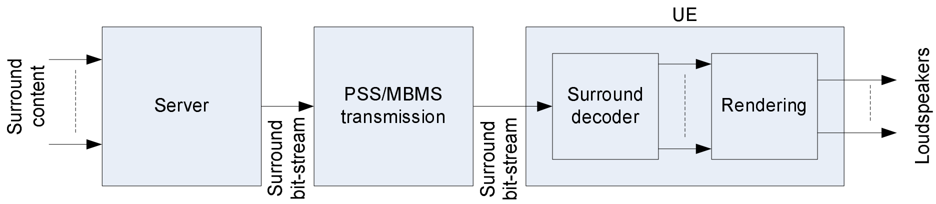
Figure 2: Signal flow for use case 1 b where binaural processing is part of the surround sound decoder

4.2 Surround sound over loudspeakers

In case the surround sound is to be played back over loudspeakers a number of scenarios can be considered. These scenarios are outlined below.

4.2.1 Decoding and rendering on a UE

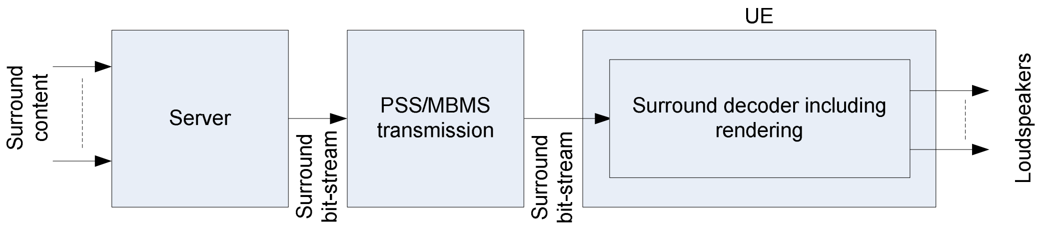
A loudspeaker scenario is illustrated in Figures 3 and 4. In this use case the surround bit-stream is first decoded in the UE to a surround sound signal. Then, in a next step a separate rendering process is applied inside the UE to map the surround sound signal onto the particular loudspeaker configuration directly connected to the UE. It is to be noted that the number of channels after surround decoding can be different from the number of channels after the rendering process.



NOTE: The resulting surround signal is input to a rendering block inside the UE that produces a representation of the surround signal for loudspeaker reproduction.

Figure 3: Signal flow for use case 2.1 a The surround bit-stream is decoded inside the UE

An alternative implementation is shown in Figure 4, where the rendering is part of the surround decoder.

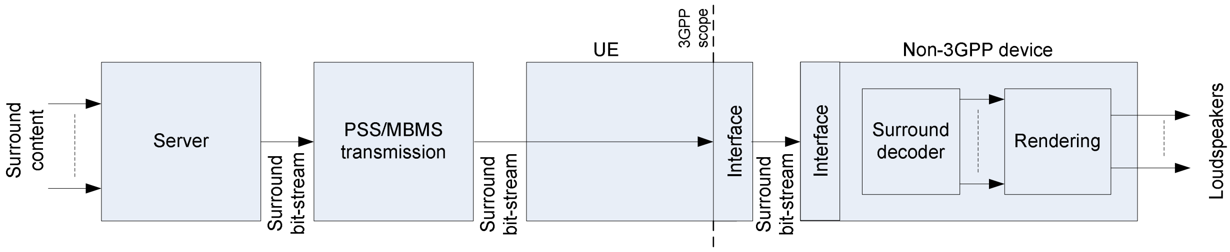


NOTE: . The surround bit-stream is decoded inside the UE while the rendering for loudspeakers is part of the surround decoding process.

Figure 4: Signal flow for use case 2.1 b

4.2.2 Decoding and rendering on a non-3GPP device connected to a UE

Another loudspeaker scenario is illustrated Figure 5. In this scenario the UE acts as an interface to a non-3GPP device. It therefore does not decode the surround bit-stream. The surround decoding and (optional) rendering is performed on the connected non-3GPP device.



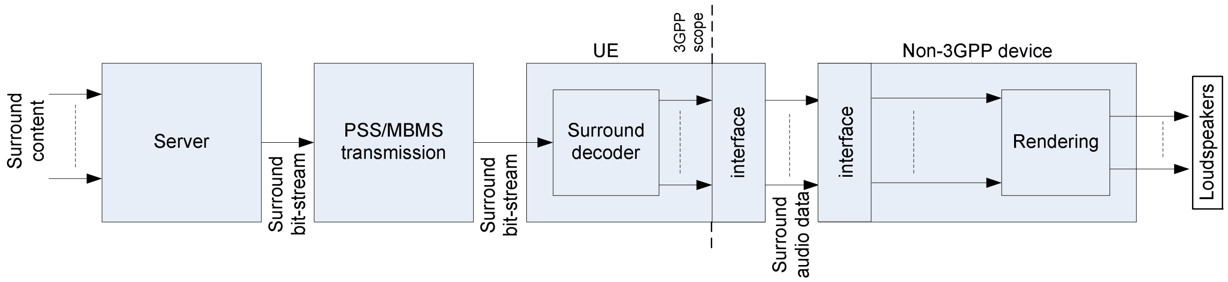
NOTE: The surround bit-stream is transported via the UE. The surround bit-stream is decoded in a non-3GPP device where also rendering takes place to produce a representation of the surround signal for loudspeaker reproduction.

Figure 5: Signal flow for use case 2.2

In this case the UE does not decode the surround sound. Instead the surround sound bit-stream is passed on to a connected device via an interface. The interface defines the transport mechanism and physical means of transporting the signal in a way that is compatible to the non-3GPP device. Therefore, the interface definition is out of the scope for this 3GPP specification.

4.2.3 Decoding on a UE and rendering on a non-3GPP device connected to a UE

Finally, a third loudspeaker scenario is illustrated in Figure 6. In this case the surround bit-stream is decoded inside the UE to multi-channel PCM. Then, via an interface provided by the UE the surround audio data is provided to a non-3GPP device where (optional) rendering takes place.



NOTE: The surround bit-stream is decoded inside the UE. The resulting surround sound audio data is transported via the UE to a non-3GPP device. The non-3GPP device renders the surround sound signal to a representation of the surround signal for loudspeaker reproduction.

Figure 6: Signal flow for use case 2.3

In this scenario the UE handles the decoding of the surround audio bit-stream. The decoded surround audio data are passed on to a connected device via an interface. The interface defines the transport mechanism and physical means of transporting the signal in a way that is compatible with the non-3GPP device. Therefore, the interface definition is out of the scope for this 3GPP specification, while the surround decoder is inside the 3GPP scope.

4.2.4 PSS/MBMS delivery methods

The PSS/MBMS delivery methods, e.g streaming, download, for the different uses cases of surround sound over 3GPP are described in particular in TS 26.234 [3] (PSS) and TS 26.346 [2] (MBMS).

The bit rates used to deliver surround sound will depend on the capabilities of the bearer, the content and the service. For instance, for streaming the surround sound codec is expected to operate at 32 kbps and upwards.

5 Design constraints

The design constraints are defined in the following subclauses.

5.1 Mono/Stereo Backwards compatibility

It is recommended that the surround sound codec should offer mono/stereo backward compatibility to existing 3GPP audio codecs.

The surround sound bit-stream should be constructed in a way that a non-surround-capable terminal can ignore the additional side information that would be used by the surround sound decoder to convert the mono or stereo core coder bit-stream into a multichannel signal, i.e. the surround sound bit-stream should be constructed in a way that a non-surround-capable terminal should be able to decode the mono or stereo backward-compatible part of the bit-stream.

When the core coder is stereo, the stereo signal contained in the surround sound bit-stream could be either an automatic stereo downmix of an initial multichannel content or an artistic stereo downmix provided. The decoder should be able to detect the presence of an artistic downmix and decode the multichannel signal on the basis of this downmix.

When the core coder is stereo the initial multichannel content will be, by default, downmixed to stereo by the encoder. Since the automatic downmix performed by the encoder may not have a sufficient quality compared to an artistic downmix, the surround encoder should be able to use as input a stereo artistic downmix and deliver the surround parameters accordingly to the decoder.

The choice for mono or stereo backward compatibility is an encoder choice, which will typically depend on the total bit rate available for coding of the multi-channel content.

5.2 Number of audio channels

5.2.1 Number of audio input channels

It is recommended that the surround sound bit-stream format supports delivery of 5.1ch content, which is the de facto standard in home entertainment industry.

The surround sound bit-stream format may support delivery of content containing more than 2 but less than 5.1 channels.

The surround sound bit-stream format may support delivery of 7.1ch content. In that case the surround sound decoder shall be capable of creating 5.1ch output from a 7.1ch bit-stream.

5.2.2 Number of audio output channels

For the purpose of this study phase, it is recommended that the surround sound decoder shall be able to provide stereo output for all surround sound bit-streams, i.e. also for a mono core coder bit-stream a stereo output should be decoded.

If the UE supports listening over loudspeakers according to Figure 6, it is recommended that the surround sound decoder shall be able to provide a multi-channel audio output for playback over multichannel loudspeakers.

If the UE supports listening over headphones according to Figure 1, it is recommended that the surround sound decoder shall be able to provide a multi-channel audio output for additional post-processing.

If the UE supports listening over headphones according to Figure 2, it is recommended that the surround sound decoder shall be able to provide a stereo output and binaural stereo output, i.e. a virtual surround output for playback over headphones.

If the UE supports listening over stereo loudspeakers according to Figure 4 with two loudspeakers, it is recommended that the surround sound decoder shall be able to provide a stereo output and a stereo output with spatial effects, i.e. a virtual surround output for playback over stereo loudspeakers.

5.3 Sampling frequency

The maximum allowed sampling rate is 48 kHz.

5.4 Bit rates

It is recommended that the surround sound codec shall support a minimum bit rate of 32 kbps.

The bit rate of the surround codec is defined as the total resulting bit rate needed to encode the multichannel audio signal, i.e. including any mono or stereo core codec bit-stream.

5.5 Computational complexity

This subclause applies to use cases where the decoding and/or rendering are performed on the UE.

The complexity is a relevant consideration in an on-the-go scenario where the UE relies on battery power to operate. In this scenario, playback would typically be over headphones, preferably using a binaural mode to create a spatial impression.

If discrete surround sound would be delivered, the decoder complexity would be approximately 2.5x the complexity of a stereo 3GPP audio decoder. Binaural processing would add on top of that. The proposed design constraint would require a lower computational complexity for the surround sound decoder. (see TR 26.936 [7] for guidance of complexity figures).

For the purpose of this study phase, it is recommended that a surround sound decoder with a binaural stereo output shall offer a lower computational complexity than 2.5x of a stereo core decoder.

An objective is to offer a minimum complexity operating mode with a reduced but acceptable quality: It is recommended that a surround sound decoder in this mode with a binaural stereo output should offer a lower computational complexity than 1.5x of a stereo core decoder.

For the purpose of this study phase, it is recommended that a surround sound decoder with a multichannel output shall offer a lower computational complexity than 3x of a stereo core decoder.

5.6 Other design constraints

If the UE supports listening over headphones according to Figure 2, the binaural surround decoder shall provide an interface to Head-Related Transfer Functions (HRTFs) and an example set of HRTFs should be provided to test the HRTF interface performance.

The decoder may be able to receive an optional input from a head tracking device to provide head-tracking.

6 Test item selection criteria

One of the important parameters in a subjective test is the selection of appropriate items. In the context of evaluating surround sound codecs, the following should be considered.

First of all, the test items should consist of multi-channel content, where a significant contribution of the audio signal is in the surround channels. No quality improvement of surround can be expected for content that could also be represented by a stereo signal. I.e. also that at least some items should be selected that contain discrete sounds from the surround channels.

Secondly, the items should be typical for and balanced over the intended application scenarios (see clause 4). This will aid justification of the surround sound codec for PSS/MBMS use cases.

The items should have the following parameters:

- Duration in the range of 7 seconds to 20 seconds

- The length of the sequences should typically not exceed 20 seconds to avoid fatiguing of listeners and to reduce the total duration of the listening test [5].

- 5.1 channels,
- 16 bit,
- 44.1 kHz.

The items should cover the genres listed in the table 2.

Table 2: List of genres considered for the item selection

Genre	Justification
Movie sound effects/dialog/trailer The items should be excerpts from movies with surround effect as well as excerpts from scenes containing speech (dialog) with surround sound. The items should also contain movie trailers.	Movies trailers and TV episodes are expected to be of major importance for mobile consumption and should be well represented in the selected test items.
Music studio production/live event The items should be excerpts from commercial music productions in studios. Also live concert recordings should be included both from popular and classical music.	Today, more than 1500 Surround music albums and over 10,000 concerts & music videos are produced with increasing numbers. Therefore, this is considered a second important category to be included in the test.
Sport event broadcast commentary/audience The items should be excerpts from live events such as sport events or other live events including commentary and audience noise.	Broadcast of live sport events is popular content. It is relevant to test whether surround sound is able to contribute to the sensation of the live atmosphere.
Radio drama The items should be excerpts from radio play productions that make use of surround sound, where speech and ambient sounds are present.	While representing only a small share of the market audio-only productions should be included in the item selection. E.g. they can make use of surround sound as means to deliver a sense of the space where the actions takes place.

The results of the item selection and the description of the test methodology are described in the test plan in Annex A.

7 Performance requirements

In this clause performance criteria are collected. First some general performance requirements are proposed. Then a set of requirements directly related to the tests specified in the test plan are proposed. Any deviation from the criteria should be justified by the performance in other tests relevant to evaluate the surround solution for 3GPP.

7.1 General requirements

It is recommended that the surround codec provides a surround experience for both loudspeaker and headphones configurations. Increased bit rate of the surround codec should be commensurate with the gain in quality of user experience.

The quality of the surround codec should be evaluated in error-free conditions and simulations of conditions as close as possible to realistic PSS/MBMS communication scenarios.

7.2 Loudspeaker requirements

Average quality is evaluated over listeners and items.

1) 64 kbps surround sound condition:

    a) For 64 kbps the audio quality should be graded on average at least as "good" on the MUSHRA scale

   b) For 64 kbps the audio quality on average should be better than the discrete 5.1 condition at 64 kbps

2) 64 kbps surround sound condition based on ITU downmix:

   a) The audio quality for the 64 kbps condition where the ITU downmix is employed should on average be comparable to the 64 kbps condition employing the default downmix.

3) 48 kbps surround sound condition:

   a) Comparing the audio quality of the 48 kbps condition to the 64 kbps condition, the audio quality for the 48 kbps condition should be commensurate to the corresponding bit rate reduction.

4) 96 kbps surround sound condition:

   a) Comparing the audio quality of the 96 kbps condition to the 64 kbps condition, the audio quality for the 96 kbps condition should be commensurate to the corresponding bit rate increase.

7.3 Binaural test

1) Integrated binaural surround sound decoder (B1) or surround sound decoder with binaural post-processing (B2):

   a) B1 should on average provide an improvement over stereo downmixing followed by binaural post-processing at the same overall bit rate

   b) B2 should on average provide an improvement over stereo downmixing followed by binaural post-processing at the same overall bit rate

7.4 Backward compatibility test

1) The surround sound codec needs to be able to employ different downmixes in the underlying compatible stereo stream (condition 2 of the loudspeaker test).

2) First, the average bit rate of the surround sound extension should be quantified at the typical bit rates.
   Then, using a stylized curve of the underlying core coder the quality degradation to the core can be approximated.
   The quality impact of including the surround sound extension to a stereo service should not exceed 20 points on a MUSHRA scale and should be as small as possible for the typical bit rate range of the surround sound codec.

7.5 Error test

The average deterioration observed at random frame error rates of 1% and 3% should not be significantly larger than the deterioration observed in similar tests using the R6 audio codec selection phase (see 3GPP TR 26.936 [7], clause B.1).

7.6 Listening test on HRTF

In general listeners should prefer the binauralized multichannel content to its stereo version produced by an ITU-R downmix [4].

8 Validation of the user benefits and feasibility through evaluation of at least one example of surround sound

8.1 Listening test over loudspeakers

Results are quite consistent between all test sites although it has been noticed that the scores of the listeners at FhG are generally lower than the scores from France Telecom, Huawei and Samsung.

The item category does not have a major influence on the results of the codecs under test.

The majority of codecs under test have been judged "Good" in quality, with no significant differences between MPS64 results and MPS ITUdmx64 results (see Figures 7 and 8), both being scored statistically better than the MPS48 and the HEAAC64. The results of those 2 last mentioned codecs are also not significantly different, meaning that the codec under test MPS48 is not significantly different from the indicative reference condition HEAAC64.

The overall quality of the MPS96 has been scored "Excellent" on average, although on the sharp border with the "Good" range, 8 points lower than the indicative reference condition HEAAC160.

On average the MPS96 condition is scored higher than both MPS64 conditions for all test sites.
For France Telecom, Huawei and Samsung test sites, the MPS96 condition scores in the excellent range and is not significantly different from the MPS64 conditions. For the FhG test site the MPS96 condition scores in the good range and is significantly better than the MPS64 conditions.

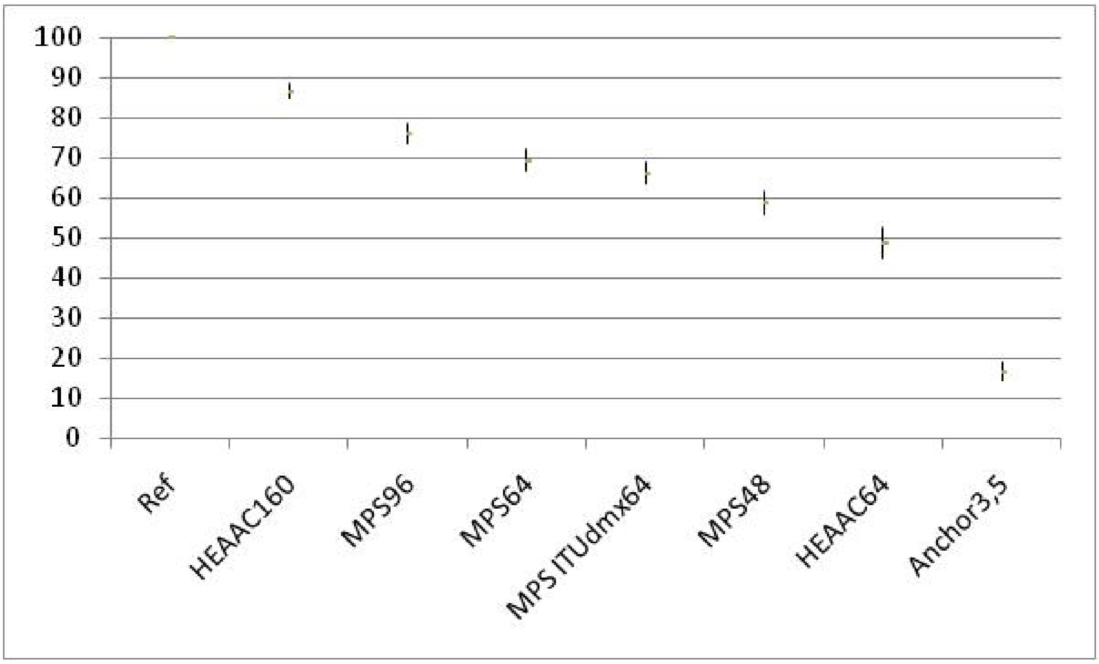


Figure 7: Global results for FhG test site

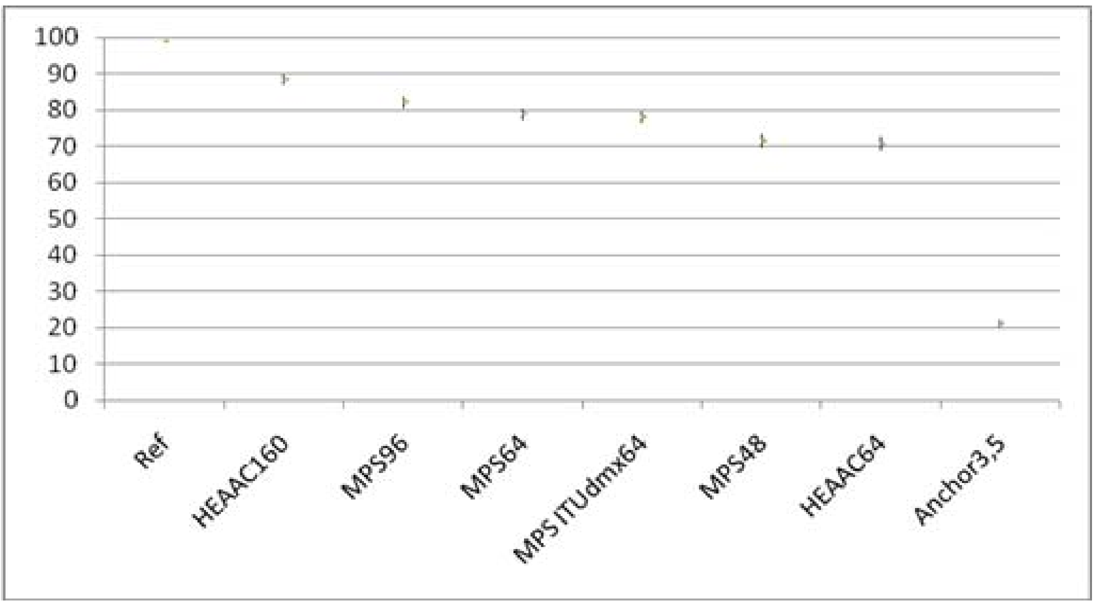


Figure 8: Results for three test sites (France Telecom, Huawei and Samsung)

Based on the results above, the codec under test, consisting of the combination of HE-AAC and MPEG Surround, met the first three performance requirements of clause 7.2. The fourth performance requirement of clause 7.2 was met for one test site (Fraunhofer), while the requirement was not met in the group of three test sites (France Telecom, Huawei and Samsung).

8.2 Listening test over headphones

This test was run twice. The first test plan used the "comparative MUSHRA" methodology . The results of this test were found to be inconclusive based on the 1st GAL report.

Therefore, a new test was designed using BS.1284 methodology. The conclusions in this clause are based on the conclusions from the 2nd GAL report.

An overview of the test conditions and pairs is provided in Table 3.

Table 3: Description of Test Conditions and Quality Comparisons

Cond	Surround-sound condition	
C1	High-bitrate surround	HE-AAC 5.1 at 320 kbps + binaural post-processing
C2	Evaluation codec	MPS 5.1 with HE-AAC stereo core codec with binaural post-processing
C3	Evaluation codec	MPS binaural decoding with HE-AAC stereo core codec
C4	Stereo downmix	HE-AAC Stereo downmix at 64 kbps + binaural post-processing
C5	Low-bitrate surround	HE-AAC 5.1 at 64 kbps + binaural post-processing
C6	High-bitrate stereo	HE-AAC Stereo downmix at 128 kbps + binaural post-processing
C7	Server side surround anchor	Binaural (post-)processing encoded with HE-AAC at 64 kbps
Comp	**Quality Comparison**	
C1-C1	Control condition: High-bitrate surround (1) vs. High-bitrate surround (1)	
C1-C6	Reference condition: High-bitrate surround (1) vs. High-bitrate stereo (6)	
C3-C4	Evaluation codec (3) vs. Stereo downmix (4)	
C3-C5	Evaluation codec (3) vs. Low-bitrate surround (5)	
C2-C4	Evaluation codec (2) vs. Stereo downmix (4)	
C7-C2	Server side surround anchor (7) vs Evaluation codec (2)	

The results of all labs averaged over all items are shown in Figure 9a.

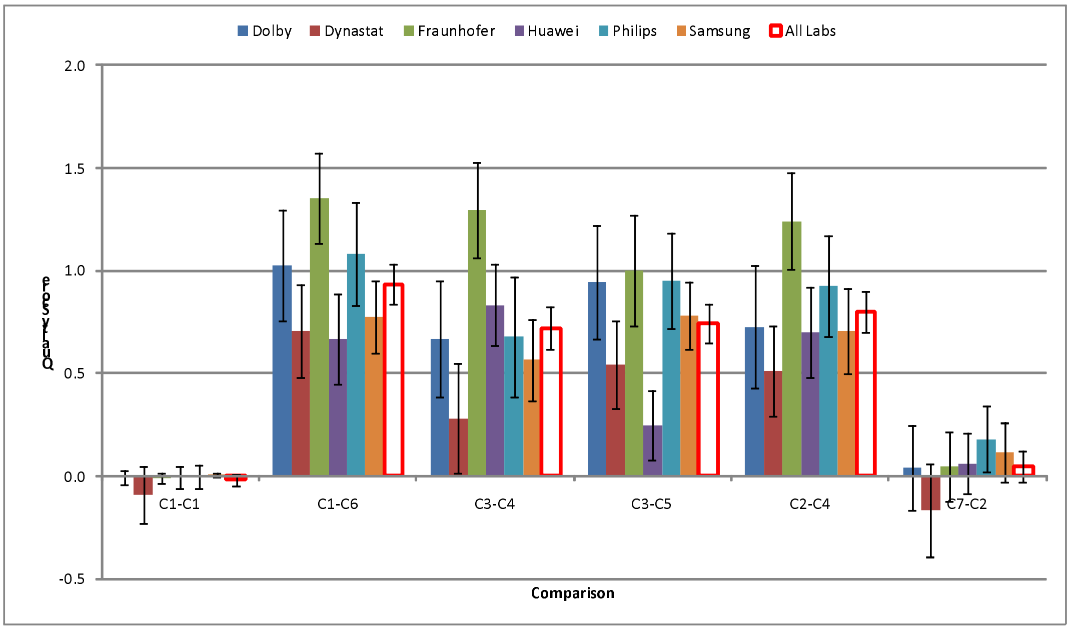


Figure 9a: Mean Scores and 95% Confidence Intervals for the Test-2 Comparisons

The results of the Global Analyses of the Test 2 data (text extracted from the GAL report in Annex A) indicated the following:

- The control comparison, C1-C1, was not significantly different from zero in any of the six Labs
- The comparison C7-C2 was not significantly different from zero in five of the six Labs
- The comparisons C1-C6, C3-C4, C3-C5, and C2-C4 were all significantly different from zero in all six Labs.
- ANOVA comparison of C3-C4 vs. C2-C4 was not significant in any of the six Labs
- ANOVA comparison of C3-C4 vs. C2-C4 was not significant across the six Labs

The patterns of scores across Labs are quite similar.

The Mean Scores and 95% Confidence Intervals for the Test-2 Comparisons shown on the full range of quality scores are shown in Figure 9b.

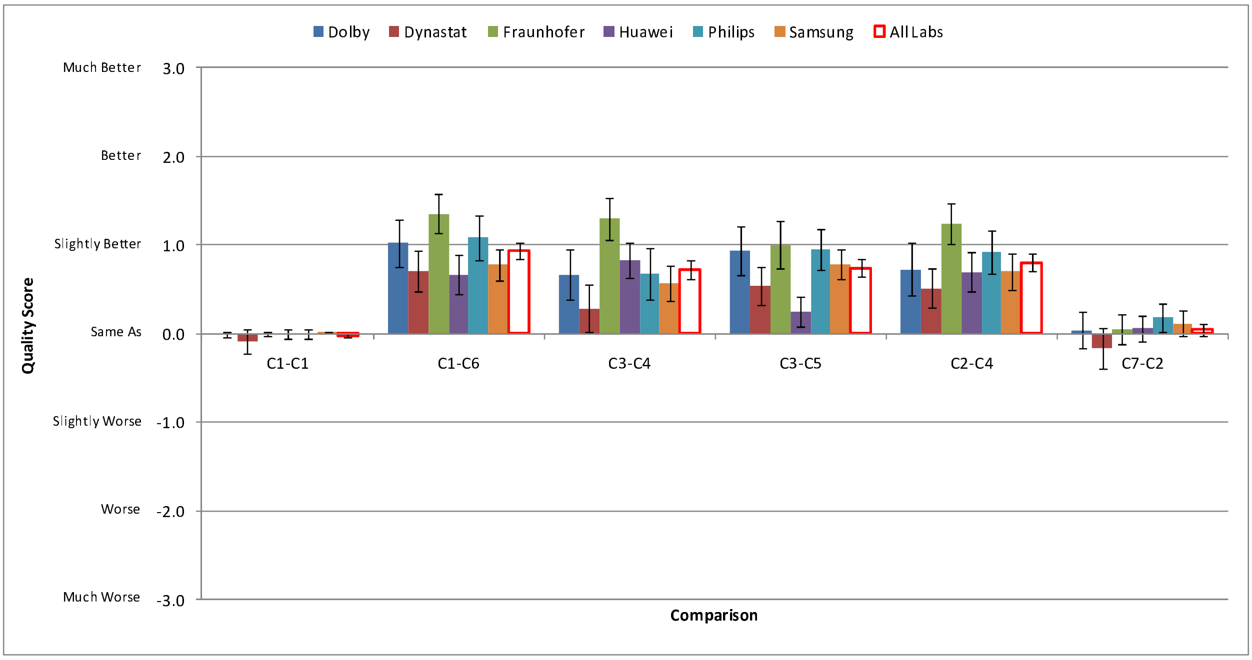


Figure 9b: Mean Scores and 95% Confidence Intervals for the Test-2 Comparisons shown on the full range of quality scores

Based on the test results, the following conclusions can be drawn.

C1-C6, this means that for all labs the high bit-rate surround (320kbps) is statistically better than the high bit-rate stereo (128 kbps) condition. According to the labels of the methodology, C1 (Surround at 320 kbps) is "slightly better" than C6 (stereo at 128 kbps).

C3-C4, this means that for all labs the evaluation codec with integrated binaural decoding is statistically better than the stereo downmix condition. According to the labels of the methodology, C3 (MPS at 64 kbps with integrated binaural decoding) is between "same as" and "slightly better" than C4 (5.1 ITU downmix HE-AAC stereo 64 kbps + binaural post-processing).

C3-C5, this means that for all labs the evaluation codec with integrated binaural decoding is statistically better than the low bit-rate discrete surround condition. According to the labels of the methodology, C3 (MPS at 64 kbps with integrated binaural decoding) is between "same as" and "slightly better" than C5 (dicrete 5.1 HE-AAC at 64 kbps).

C2-C4, this means that for all labs the evaluation codec with binaural post-processing is statistically better than the stereo downmix condition. According to the labels of the methodology, C2 (MPS + binaural postprocessing at 64 kbps) is between "same as" and "slightly better" than C4 (5.1 ITU downmix HE-AAC stereo 64 kbps + binaural post-processing).

The differences between C2-C4 and C3-C4 were statistically equivalent to zero, i.e C2 and C3 were equivalent based on the quality scale shown in Fig 9b. According to the labels of the methodology, C2 (MPS + binaural postprocessing at 64 kbps) is "same as" C3 (MPS at 64 kbps with integrated binaural decoding).

C7-C2, this means that for all but one lab the evaluation codec with binaural post-processing is statistically equivalent to the server-side surround anchor. According to the labels of the methodology, C7 (server-side binaural conversion + HE-AAC at 64 kbps) is "same as" C2 (MPS + binaural postprocessing at 64 kbps).

8.3 Backward compatibility

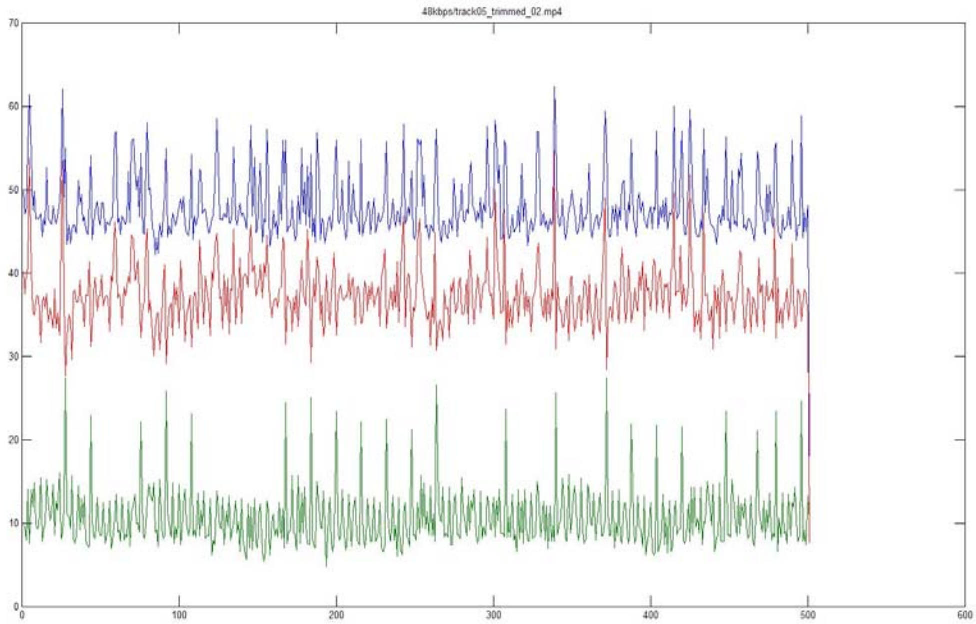
According to Test 3 of the test plan the quality impact due to the inclusion of a surround sound extension, which takes away a certain bit rate from the underlying stereo core coder, has been determined for the surround sound codec configurations described in the test plan. Table 4 provides an overview of the tested configurations, all consisting of combinations of HE-AAC with MPEG Surround.

Table 4: Overview of tested surround sound configurations

Configuration	Overall bit rate (kbps)	HE-AAC bit rate (kbps)	MPEG Surround bit rate (kbps)
HE-AAC/MPS @ 48kbps	48.0	39.8	8.2
HE-AAC/MPS @ 64kbps	64.0	55.8	8.2
HE-AAC/MPS @ 96kbps	96.0	66.7	29.3 (see Note)

In order to establish an estimate of the aforementioned quality impact, a bit rate versus quality curve is required for the HE-AAC codec. This bit rate versus quality curve is primarily established by taking MUSHRA measurement points from the 3GPP audio codec selection tests [7] at 18, 24, 32 and 48 kbps. Additionally, an estimate for the quality at 128 kbps [6] is provided. However, this bit rate was not formally evaluated. An estimated bit rate versus quality curve is then established by applying a piecewise cubic interpolation through the average MUSHRA scores. This is visualized in Figure 10.

NOTE: For the HE-AAC/MPS @ 96 kbps configuration so called 'residual coding' is employed. This means that in addition to spatial parameter data, additional waveform coded data is employed to further improve the multi-channel audio quality.



NOTE: The 3GPP high rate selection results are indicated as mean and 95 % confidence intervals. The additional data point at 128 kbps is indicated by a circle

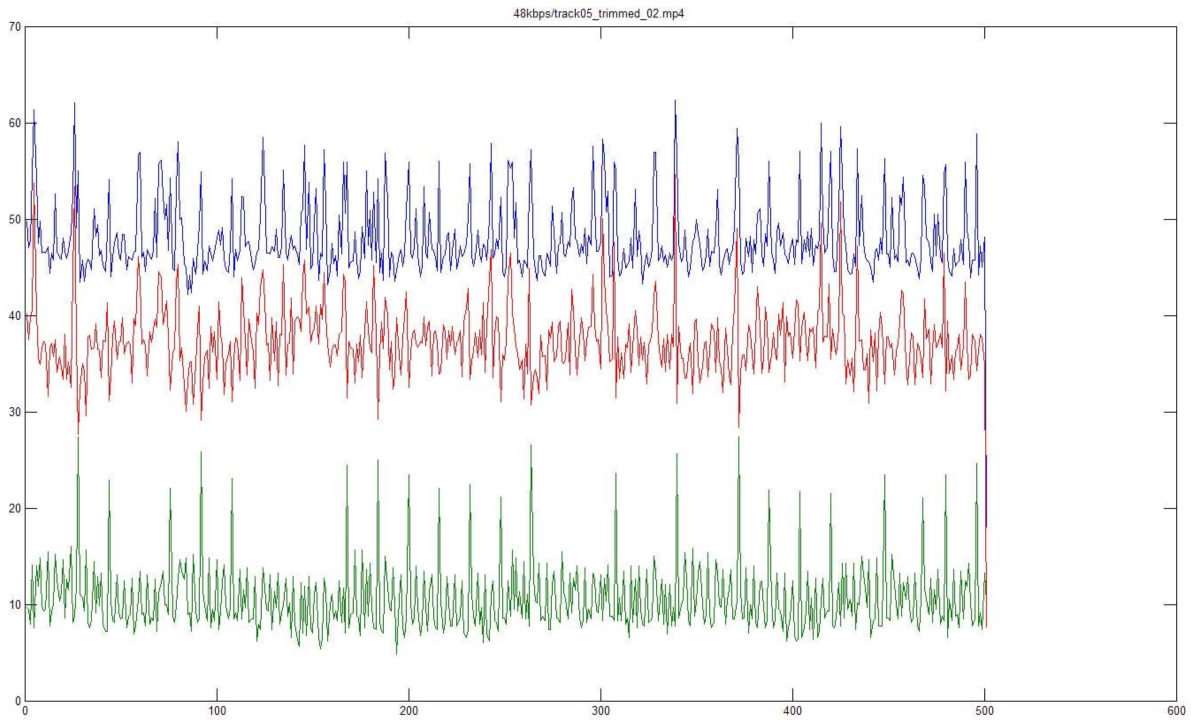
Figure 10: Graphical representation of bit-rate versus quality curve for HE-AAC codec

Using the information from Table 4, an estimate is made for the quality reduction by comparing the audio quality at the overall bit rate (e.g. 48 kbps), which could be used if no surround sound extension would be employed, versus the quality at the bit rate consumed by the stereo core (e.g. 39.8 kbps). This process is also shown in Figure 10, where the horizontal lines denote the bit rate differences and the vertical lines the quality differences. The estimated quality differences are provided in Table 5.

Table 5: Quality estimate of stereo core for 48, 64 and 96 kbps configurations in case of HE-AAC or HE-AAC/MPEG Surround combination

Configuration	MUSHRA score for HE-AAC only	MUSHRA score for HE-AAC/MPS	MUSHRA quality difference
48 kbps	82.0	79.8	-2.2
64 kbps	85.7	83.9	-1.8
96 kbps	91.2	86.3	-4.9

It should be noted that the above values are average bit rate values. The actual instantaneous bit rate spent for each frame is variable, as it is in the underlying HE-AAC stereo coder. The bit rate distribution in Figure 11 shows that the variability of the MPEG surround data rate is basically synchronous to the variability of the HE-AAC stereo core coder.



NOTE: In blue the total is depicted; in red the stereo core bit rate is depicted and in green the share of the surround bit rate is shown.

Figure 11: Example of the instantaneous bit rate distribution of an MPS bit-stream

Based on the results presented above, the codec under test, consisting of the combination of HE-AAC and MPEG Surround meets the performance requirements of clause 7.4.

8.4 Test under errors conditions

8.4.1 Results with interleaver

Results are quite consistent between test sites although it has been noticed that the scores of the listeners at FhG are generally lower than the scores from Philips.

The item category does not have a major influence on the results of the codecs under test.

All 8 codecs under test have been judged "Good" in quality.

Results of the Student test show that whatever the frame error rate value (random or interleaved bursty), there is no significant difference between the MPS with HE-AAC stereo core 64 kbps decoded in binaural mode and the MPS 5.1 with HE-AAC stereo core 64 kbps with binaural post-processing.

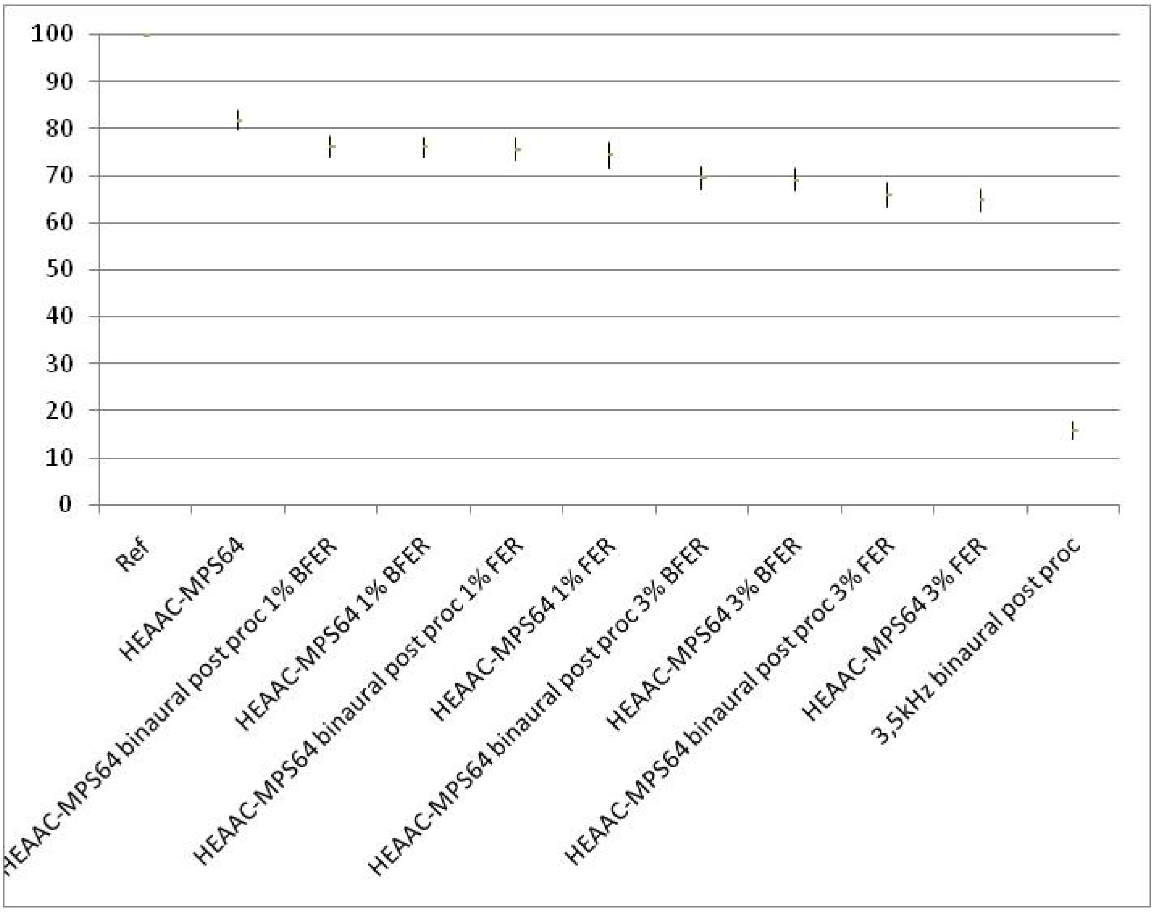


Figure 12: Test 4 results from Fraunhofer

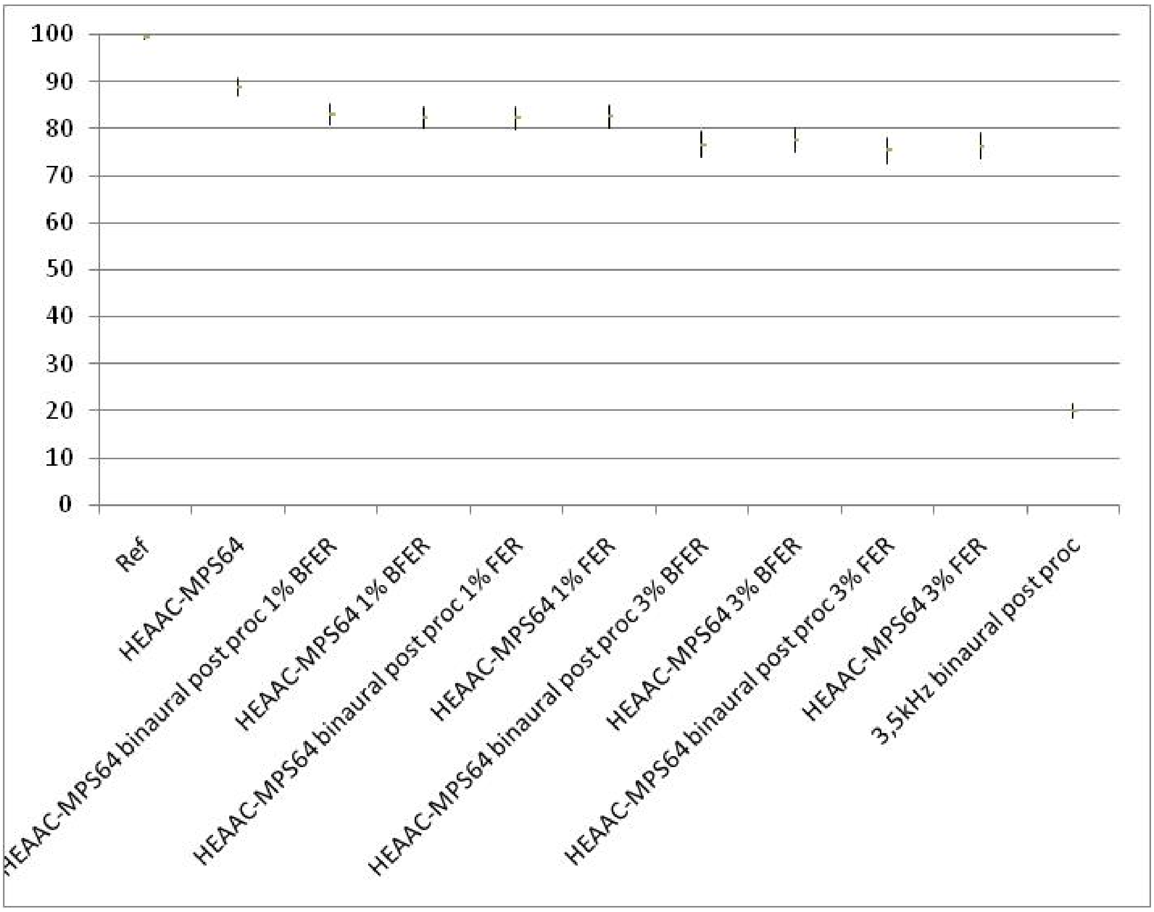


Figure 13: Test 4 results from Philips

8.4.2 Results without interleaver

Results are very consistent between test sites.

The category of items has not a major influence on the codecs under test results, except for the "Radio" category at the 1% bursty error frame rate.

For a given pattern of frame rate error, there is no difference between MPS with HE-AAC stereo core 64 kbps, decoded in binaural mode and MPS 5.1 with HE-AAC stereo core 64 kbps, with binaural post-processing.

For a given frame rate error value, the bursty pattern lowers the perceived quality comparing to a random pattern.

Anyhow, the reference codec HEAAC-MPS64 remains the highest scored in quality excepting the hidden reference.

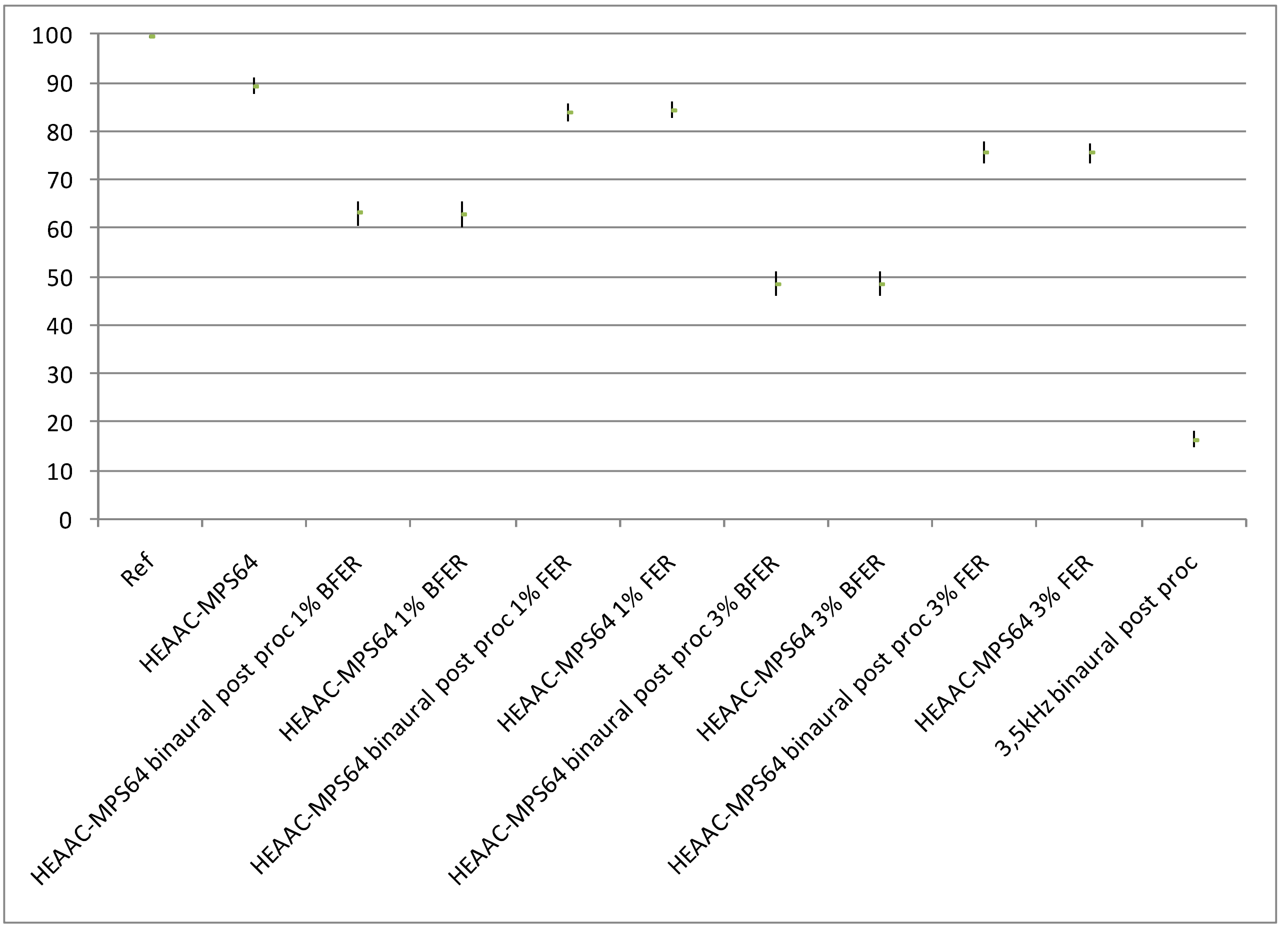


Figure 14: Global results for all test sites for the test without interleaver

In Annex B1 of TR26.936 test results are presented for the high rate tests of the R6 audio codec selection procedure. From these results an estimate of the quality impact of random frame errors can be derived. The results for the tested codecs as well as the corresponding deterioration are given below.

Table 6: Average ratings and deterioration for the R6 audio codec selection procedure

Tests	Operating condition	AAC+ (HE-AAC)	Deterioration	CT (HE-AAC v2)	Deterioration
1	32 kbps, stereo	75.8	-	84.9	-
3-1	32 kbps, stereo, 1% FER	66.2	9.6	72.9	12.0
3-2	32 kbps, stereo, 3% FER	56.3	19.5	62.3	22.6

Table 7: Average ratings and deterioration for the codecs in test 4

Tests	Operating condition	HE-AAC+MPS 64 binaural	Deterioration	HE-AAC+MPS64 5.1 + binaural post-proseccing	Deterioration
4	64 kbps, stereo	89	-	89	-
4	64 kbps, stereo, 1% FER	84	5	85	4
4	64 kbps, stereo, 3% FER	76	13	75	14

By approximation, this means that both codecs tested provide a degradation in the order of 10 MUSHRA points for 1% random FER and 20 MUSHRA points for 3% random FER. Comparing this to the system under test, the combination of HE-AAC and MPEG Surround, we see that both test sites show a slightly smaller degradation for the 1% and 3% random FER conditions. This means that the codec under test, HE AAC combined with MPEG Surround, would meet the proposed performance requirement of clause 7.5.

8.5 Test on HRTFs

The test was conducted in two listening labs. In the first listening lab (France Telecom), two headphones (Stax and Sony) were used while in the second listening lab (Huawei) an open headset was used.

HRTF set A is the KEMAR set. The details of HRTF sets B, C and D were not disclosed.

The results obtained by Stax and Sony show more similarities than between Stax and open headset (see Figures 15 and 16). The difference between Stax and Sony headphones has no influence on the perceived quality of the tested conditions.

The category of items does not have a major influence on the conditions under test results. The results of subjective listening tests from both labs show a significant dependency on the specific HRTF set.

HRTF set A is statistically worse than the stereo reference. HRTF set A has average scores in the range between "similar to REF" to "slightly worse than REF" for one listening lab (Huawei) and has average scores in the range between "slightly worse than REF" to "worse than REF" for the other listening lab (France Telecom). For one test lab (France Telecom), HRTF set A is worse than the mono downmix anchor, which is not the case for the other test lab (Huawei); indeed, the average scores for HRTF set A were significantly different for the two test sites.

HRTF set B is statistically worse than the stereo reference. HRTF set B has average scores in the range between "similar to REF" to "slightly worse than REF" for the two listening labs.

HRTF Set C is statistically better than the stereo reference for the two listening labs, according to the labels of the employed methodology. HRTF set C has average scores in the range between "similar to REF" to "slightly better than REF",

HRTF set D is statistically worse than the stereo reference. HRTF set D has average scores in the range between "similar to REF" to "slightly worse than REF" for one listening lab (Huawei) and has average scores in the range between "slightly worse than REF" and "worse than REF" for the other listening lab (France Telecom).

The Mono downmix is statistically worse than the reference. The Mono downmix has average scores in the range between "slightly worse than REF" and "worse than REF" for the two listening labs.

It was observed that the mono anchor condition in one listening lab (France Telecom) results did not show the lowest scores in this test, as it is the case for the other listening lab (Huawei).

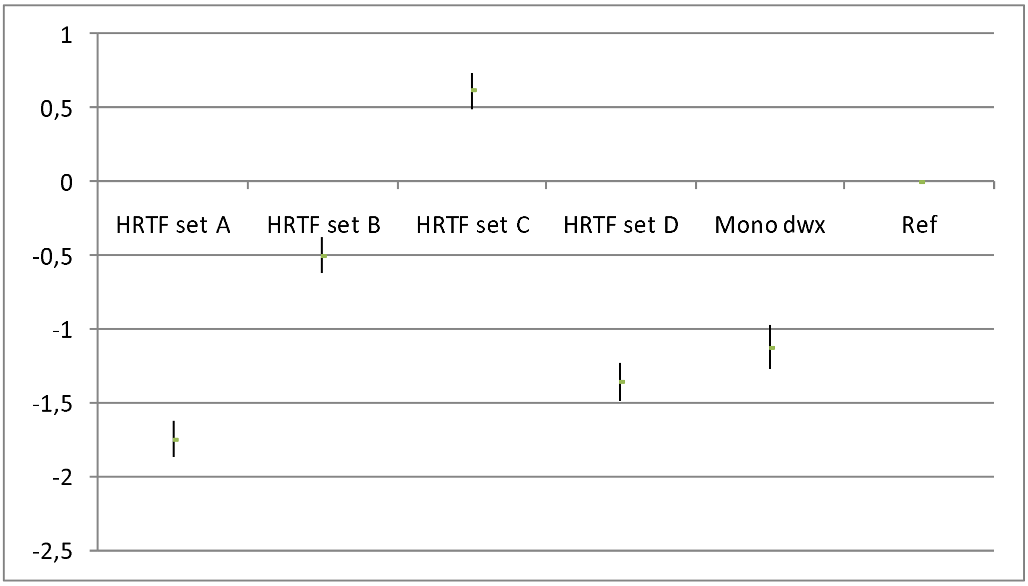


Figure 15: global results for Stax and Sony headphones

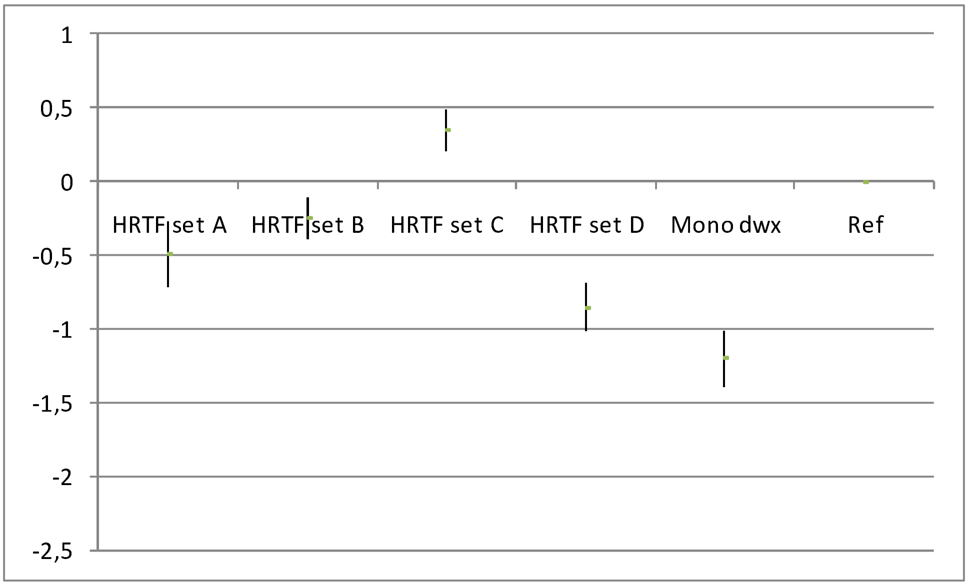


Figure 16: HW global results

9 Conclusion

In the framework of the Surround Sound Study Item, use cases were identified and investigated in the context of 3GPP services PSS and MBMS. The use cases relate both to loudspeaker and headphone reproduction. Based on those use cases a set of design constraints and performance requirements have been derived. Test methods were developed to conduct extensive subjective listening tests in order to evaluate the benefits of surround sound over stereo in the typical use cases. An example surround sound codec was used for the testing; it consists of the combination of HE-AAC and MPEG Surround.

Overall it can be concluded that the study has demonstrated that surround sound techniques can represent a user benefit in certain use cases as detailed in this report. Furthermore, the combination of HE-AAC and MPEG Surround has indicated that such user benefit can be delivered at low bitrates, in the context of existing services, and it is possible for surround sound capable terminals to include this technology. As noted in this report, some aspects around use cases, performance requirements and design constraints as well as details of the test methodology may require further work. These aspects should be considered if 3GPP decides to pursue this topic further.

Annex A: Test plans and global analysis reports

Annex A contains the following documents attached to this TR in electronic form:

Tdoc S4-091004: "Test Plan Study on Surround Sound Version 1.0"

Tdoc S4-100705: "Test Plan Study on Surround Sound - Headphones test - Version 1.0"

Tdoc S4-100347: "Revised Global Analysis Report for Study Item on Surround Sound"

Tdoc S4-100723: "Dynastat Global Analysis Report for Test 2 of the PSS/MBMS Surround Sound Study Item"

Annex B:
Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2010-03	47	SP-100029			Presented to TSG SA#47 for information		1.0.0
2011-03	51	SP-110049			Presented to TSG SA#51 for approval	1.0.0	2.0.0
					Approved at TSG SA#51	2.0.0	10.0.0

History

Document history		
V10.0.0	April 2011	Publication