INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

ORGANISATION INTERNATIONALE DE NORMALISATION

ISO/IEC JTC 1/SC 29/WG 11

CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC 1/SC 29/WG 11 **N8851**

January 2007, Marrakech, Morocco

Source: **Audio Subgroup**
Title: **Report on MPEG Surround Verification Test**
Status: **Approved**

Summary

This document reports on the verification test of the MPEG-D "MPEG Surround" audio coding technology. By considering three test scenarios, the performance of MPEG Surround using AAC, HE-AAC and MPEG-1 Layer II as core coders has been evaluated with respect to discrete multi-channel AAC, HE-AAC and matrix surround coding. In addition, three specific MPEG Surround decoding modes have been tested: binaural decoding, enhanced matrix mode and low power decoding.

In summary, the test results show that MPEG Surround offers superior compression efficiency to existing technology at lower bitrates, while scaling to the high quality of discrete systems at higher bitrates and inherently providing a backwards compatible downmix. Furthermore, the enhanced matrix mode of MPEG Surround operating without any side-information performs significantly better than legacy matrix based systems. Finally, MPEG surround offers several binaural decoding modes of excellent quality that enable a high quality surround sound experience over headphones on portable devices.

Table of Contents

1 Introduction 2

2 Codecs under test 3
2.1 Introduction 3
2.2 DVB oriented test-case 3
2.3 Music-store / portable player oriented 5
2.4 Technology driven tests 7

3 Test material 8

4 Test Centers 8

5 Test Results 9
5.1 Introduction 9
5.2 DVB oriented test-case 9
5.3 Music-store / portable player oriented 12
5.4 Technology driven test 15

6 Conclusions 17

7 Acknowledgements 17

8 References 17

A.1 Testing Schedule for the verification test 19

A.2 Pre-selection process 20
A.2.1 Collection process 20
A.2.2 Pre-selection process 20
A.2.3 Test items 21

A.3 Test methodology 23
A.3.1 MUSHRA 23
A.3.2 Training phase 23

A.4 Codec settings 25

A.5 Statistical Analysis 27
A.5.1 General 27
A.5.2 Post-screening 27

A.6 Detailed Results all test sites 29

A.7 Detailed Results individual test sites 34

1 Introduction

MPEG has issued several successful standards. The International Standards ISO/IEC 11172-3 (MPEG-1 Audio), ISO/IEC 13818-3 (MPEG-2 Audio) and 13818-7 (MPEG-2 Advanced Audio Coding, AAC) were issued in 1992, 1994, and 1997, respectively. The International Standards ISO/IEC 14496-3 (MPEG-4 Audio Version 1) and ISO/IEC 14496-3 / AMD1 (MPEG-4 Audio Version 2) were issued in mid-1999 and early 2000 respectively. In 2003 and 2004, MPEG subsequently amended the MPEG-4 specification with the parametric extensions Spectral Band Replication (SBR) and Parametric coding for high quality audio, the later also comprising Parametric Stereo (PS). These specifications have resulted in the standardization of the ISO/IEC 14496-3 HE-AAC and ISO/IEC 14496-3 HE-AAC v2 profiles respectively.

In a further step to pursue efficient parameterization of a multi-channel audio signal, at the 68th MPEG meeting in Munich 2004, MPEG issued a call on spatial audio coding. Evidence submitted in response to the call was extensively evaluated at the 70th MPEG meeting, in Palma de Mallorca. From this test the two best submissions showed complementary performance, and it was decided to merge these technologies. This resulted in a reference model (RM0) that was finalized at the 72nd MPEG meeting in Busan, Korea 2005 and

© ISO/IEC 2007 – All rights reserved

that served as the basis of further development work within WG11. The technical work on MPEG Surround was finalized by the 77th MPEG meeting in Klagenfurt, and was approved at the 79th MPEG meeting in Marrakech in the final ballot.

This document reports on the final verification test of the MPEG standardization effort on the MPEG Surround spatial audio coding technology.

The following sections outline codecs under test (Section 2), test-items (Section 3), participating test centers (Section 4), and results from the listening (Section 5).

The time-line and responsibilities of the verification test are given in Appendix A.1. Appendix A.2 summarizes the item collection and pre-selection process. In Appendix A.3 the test methodology used for the verification and the test details are outlined. Appendix 0 reports on the specific codec settings applied for generating the test material. In Appendix A.5 the statistical analysis is presented. Finally in Appendices 0 and A.7 the detailed results on a per item basis and for the individual test sites are provided.

2 Codecs under test

2.1 Introduction

In order to be able to provide relevant test data to the industry interested in the new technology, the main part of the verification testing was designed as an "application driven test". This means that in the design of the tests, two use-cases relevant to industry were considered, a DVB oriented use-case, and a music-store / portable player use-case. For each test-case a multi-channel test was performed and a corresponding downmix test or binaural decoding test. In addition a technology driven test is included that evaluates the difference between High Quality (HQ) and Low Power (LP) decoding modes of MPEG Surround. MPEG Surround LP decoding operates at significant lower computational complexity than MPEG Surround high quality decoding.

Note that the low power decoder, which is present in the binaural phase of Test 2 and in Test 3, is indicated by the “_LP” suffix in the codec name.

As specified in the MUSHRA test methodology, a hidden reference and one band-limited version of the reference were included as anchors and references in the tests.

The bitrates and configurations for the coders under test are provided in Annex 0

2.2 DVB oriented test-case

The test conditions presented in Table 1 strive to compare MPEG Surround in DVB-like applications with other potential MPEG Audio solutions for that space. Two aspects are covered:

1. Establishing a new service, for which a stereo backwards compatible part at highest quality per bit is desired.
2. Extending an existing service with surround sound in a backwards compatible way.

For the first scenario, MPEG-4 High Efficiency AAC in combination with MPEG Surround is compared with discrete multi-channel MPEG -4 HE AAC.

For the second scenario, MPEG-1 Layer 2 (as used in DVB) is extended with MPEG Surround.

© ISO/IEC 2007 – All rights reserved

Table 1 - Codecs in multi-channel test 1

Codec ID	Core Coder	Description	Total Bitrate (kbps)
HE-AAC_MPS_A	HE-AAC	High bitrate case, HE-AAC + MPEG Surround	160
HE-AAC_MPS_B (Note 1)	HE-AAC	Low bitrate case, HE-AAC + MPEG Surround	64
L2_MPS	Layer-2	High bitrate case, Layer-2 + MPEG Surround in Matrixed Downmix Mode	256
L2_DPL2	Layer-2	High bitrate case, Layer-2 + DPL2 encoding and decoding	256
HE-AAC _MC_A	HE-AAC	High bitrate reference case, HE-AAC discrete	160
HE-AAC_MC_B (Note 1)	HE-AAC	Low bitrate reference case, HE-AAC discrete	64
REF	-	Hidden reference (5.1 original)	-
LP35	-	Low pass anchor (based on 5.1 original)	-

Note 1: Has been used for the encoding/decoding for the pre-selection test.

An additional downmix test, using the operating points provided in Table 2, strives to assess the quality of the stereo backwards compatible part of "Multi-channel test 1". This test was done using loudspeakers.

Table 2 - Codecs in downmix test

Codec ID	Core coder	Description
HE-AAC _MPS_A_BC	HE-AAC	Backward compatible downmix of HE-AAC _MPS_A.
HE-AAC _MC_A_DMX	HE-AAC	ITU downmix of HE-AAC _MC_A.
REF	-	Hidden reference (ITU downmix of 5.1 original).
LP35	-	Low pass anchor (low filtered version of REF).

Figure 1 provides an illustration of the test conditions.

© ISO/IEC 2007 – All rights reserved

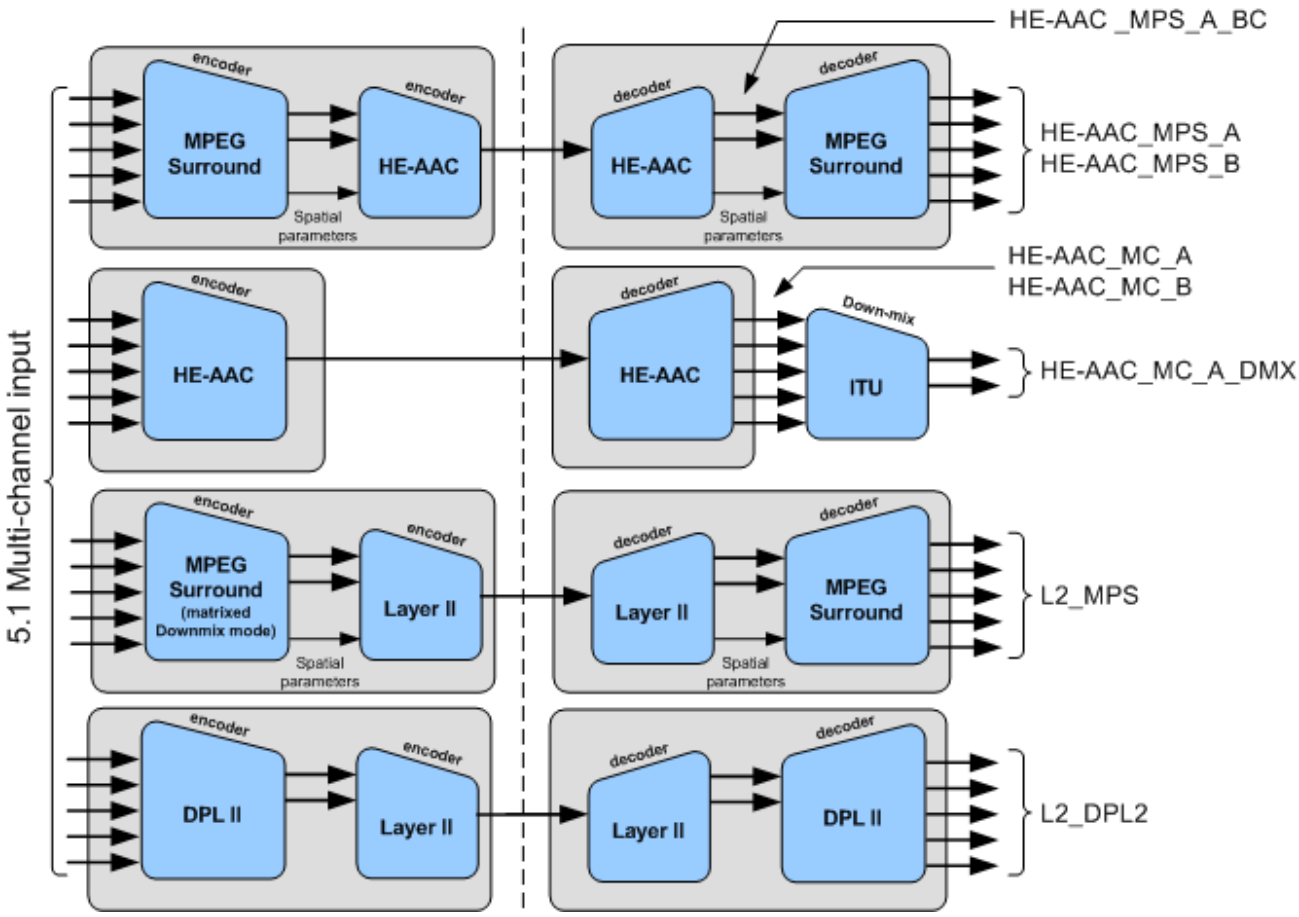

Figure 1 – Illustration of the test conditions for multi-channel test 1 and the downmix test.

2.3 Music-store / portable player oriented

The test conditions presented in Table 3 strive to test the performance of MPEG Surround in "music-store-like" applications, envisioning a scenario where an online music store selling stereo material using AAC at 160kbps, extends the service to provide multi-channel content in a backwards compatible way by additionally providing low bitrate MPEG Surround data. Hence, a consumer can play the backwards compatible part on his/her legacy player, while enjoying a multi-channel experience when playing the content at home using an MPEG Surround enabled multi-channel set-up.

Table 3 - Codecs in multi-channel test 2

Codec ID	Core Coder	Description	Core Bitrate (kbps)
AAC_MPS	AAC	High bitrate case, AAC + MPS in Matrixed Downmix Mode.	160
AAC_MPS_MTX	AAC	AAC_MPS decoded with MPS in Enhanced Matrix Mode.	160
AAC_DPL2	AAC	High bitrate reference case, DPL2 downmix encoded with AAC stereo. Decoded with DPL2	160
REF	-	Hidden reference (5.1 original).	-
LP35	-	Low pass anchor (based on 5.1 original).	-

© ISO/IEC 2007 – All rights reserved

Figure 2 provides an illustration of the test conditions.

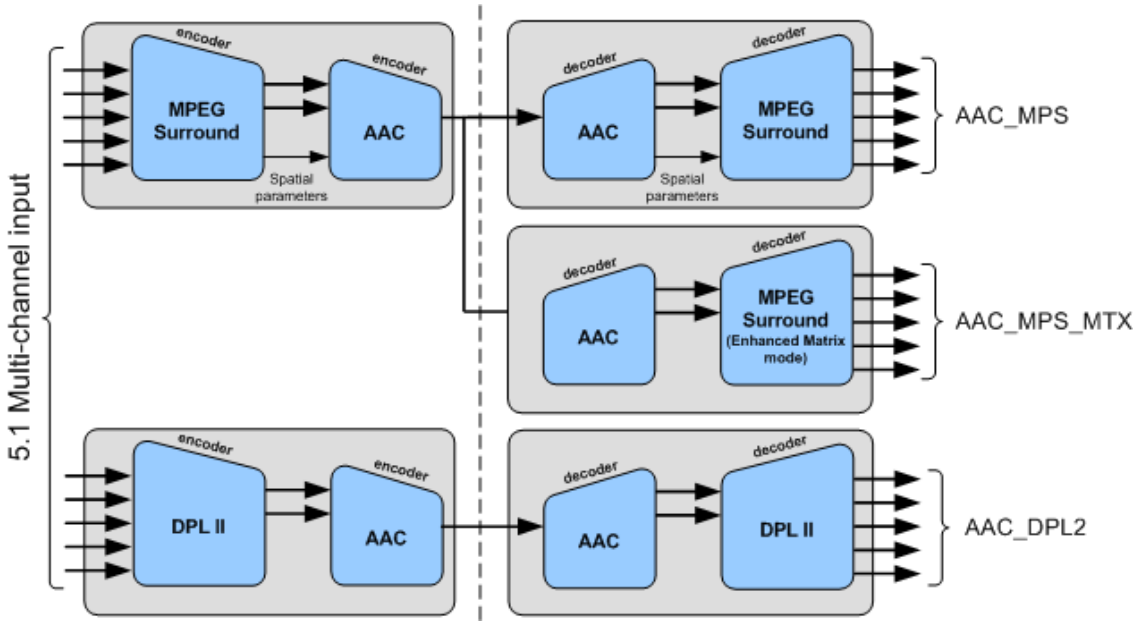

Figure 2 – Illustration of the test conditions for multi-channel test 2.

An additional binaural test, with test conditions shown in Table 4, strives to assess the quality of the binaural decoding capability of MPEG Surround, particularly for the music-store use case in "Multi-channel test 2". Two cases were tested:

- High quality binaural decoding for stereo head-phone listening at home
- Low Power binaural decoding for a portable device

Note that in binaural decoding mode, no residual is decoded.

Table 4 - Codecs in binaural test

Codec ID	Core Coder	Description	Core Bitrate (kbps)
AAC_MPS _BIN	AAC	Binaural decoding (by multi-slot convolution) of AAC_MPS.	160
AAC_MPS_LP_BIN_LP	AAC	Low power decoding (LP MPEG Surround and single-slot parametric HRTF) of AAC_MPS.	160
AAC_MPS_BIN_R	AAC	HRTFs applied to AAC_MPS multi-channel output.	160
REF	-	Hidden reference: HRTFs applied to 5.1 original.	-
LP35	-	Low pass anchor (based on signal above).	-

Although for this use case in-ear headphones (e.g. “ear buds”) typically would be used, In order to test for the most critical conditions, professional headphones were used in the subjective test. Figure 3 provides an illustration of the test conditions.

© ISO/IEC 2007 – All rights reserved

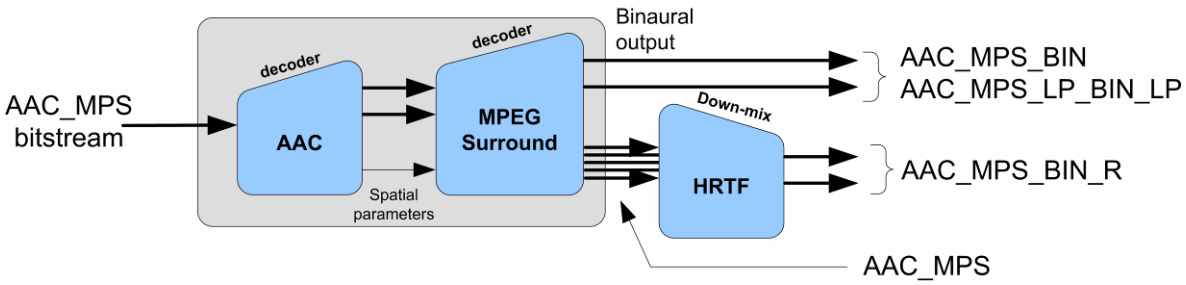

Figure 3 – Illustration of the binaural test.

2.4 Technology driven tests

Extensive technology driven testing has been performed during the development of MPEG Surround. One additional test was carried out during the verification testing in order to evaluate the difference between High Quality (HQ) and Low Power (LP) MPEG Surround decoders. In the case an MPEG Surround bitstream contains a residual component, a low power MPEG Surround decoder will only decode the residual up to a limited bandwidth, whereas the high quality MPEG Surround decoder decodes the full bandwidth of the residual.

The codecs under test in this test effort are shown in Table 5, and it is clear that the test is a derivative of multi-channel test 1.

Table 5 - Codecs in multi-channel test 3

Codec ID	Core Coder	Description	Total Bitrate (kbps)
HE-AAC_MPS_A	HE-AAC	High bitrate case, HE-AAC + MPEG Surround	160
HE-AAC_MPS_B	HE-AAC	Low bitrate case, HE-AAC + MPEG Surround	64
HE-AAC _MPS_A_LP	HE-AAC	Low power decoding of HE-AAC _MPS_A.	160
HE-AAC _MPS_B_LP	HE-AAC	Low power decoding of HE-AAC _MPS_B.	64
REF	-	Hidden reference (5.1 original).	-
LP35	-	Low pass anchor (based on 5.1 original).	-

Figure 4 provides an illustration of the test conditions.

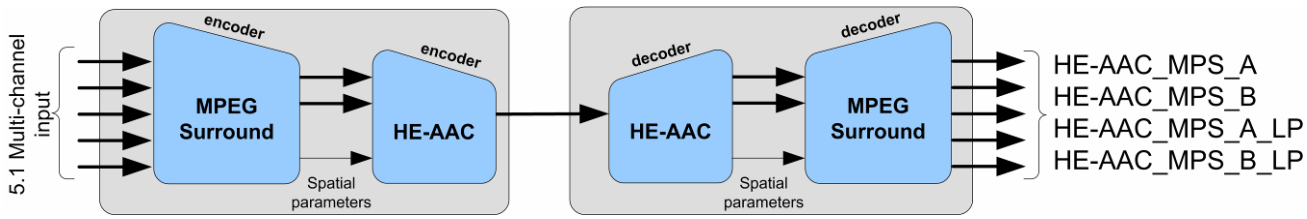

Figure 4 – Illustration of the technology driven test.

© ISO/IEC 2007 – All rights reserved

3 Test material

For the MUSHRA tests, items were selected from 159 potential candidate items by a selection panel at France Télécom R&D. The pre-selection procedure is outlined in detail in appendix A.2. The ten items listed in Table 6 were considered to contain both critical and typical items for all of the systems under test.

Table 6 – Items to be used for each test.

Label	Item name	Genre	Duration (s)	LFE
Amb1	Station_Atmo_6ch	Ambience	20	
Amb2	tennis	Ambience	24	
jaz1	tower1	Jazz	17	
Mod1	bonobo_1	Movies Drama	16	
Mod2	elliot2	Movies Drama	21	
orc1	ravel1c	Orchestra	19	
orc2	violin2	Orchestra	21	
pop1	Lavilliers_14	Pop (Lavilliers)	22	x
pop2	thalheim4	Pop	19	
pop3	SantaCruz_09a	Pop (SantaCruz)	19	

4 Test Centers

The following test sites participated in the tests. Table entries indicate the number of subjects sites provided for each test.

Table 7 – Overview of test centers and the number of subjects that participated in the various tests.

Test site	Subjects in test				
	Multi-channel test 1	Downmix test	Multi-channel test 2	Binaural test	Technology driven test
Thomson Hannover	12	-	12	-	-
France Télécom R&D	-	-	-	11	-
Coding Technologies	11	-	11	-	-
LG	-	10	-	10	-
ETRI	8	-	-	9	-
Panasonic	-	8	-	-	8
FhG-IIS	10	-	10	-	-
Philips	-	-	8	-	10
Total number of subjects	**41**	**18**	**41**	**30**	**18**

© ISO/IEC 2007 – All rights reserved

5 Test Results

5.1 Introduction

A statistical analysis and post screening (see Appendix A.5) was done on the listening test data. The subsequent plots display the mean values (horizontal tick) and 95% confidence intervals (vertical tick) averaged over all items for every coding scheme. Detailed plots of performance of codecs on a per-item basis are given in Appendix 0.

The results have been pooled over the various test sites. In order to provide a strict statistical analysis, the results from the individual test-sites are given in Appendix A.7.

Post screening has been applied to the data in order to remove subjects that did not score consistently. The procedure for post screening is provided in Annex A.5.2.

5.2 DVB oriented test-case

As outlined above, the DVB oriented test-case comprises multi-channel test 1 and downmix test 1. The plot in Figure 5 shows the results of the subjective test of multi-channel test 1 in which data from all test sites is pooled.

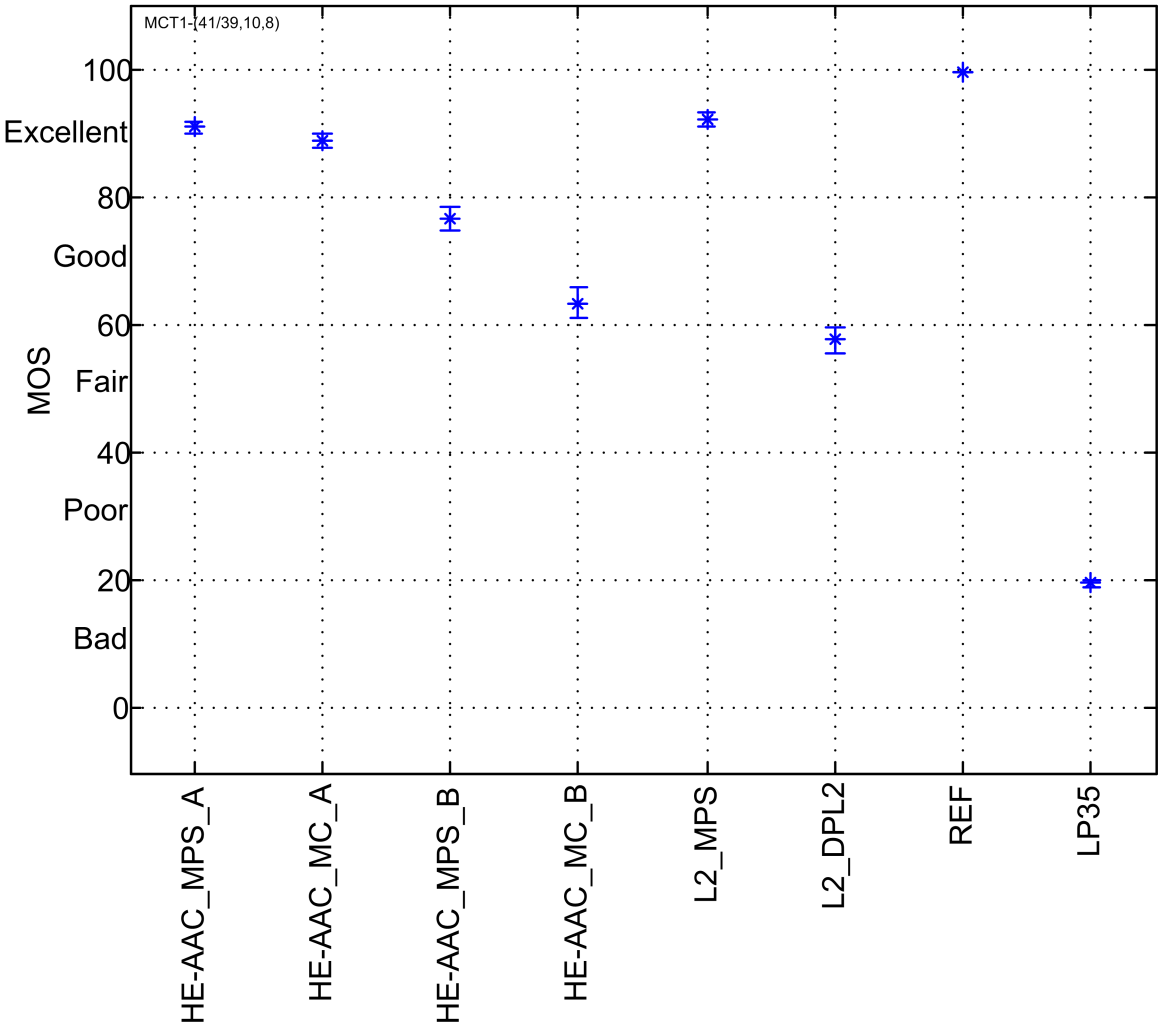

Figure 5 – Results for multi-channel test 1

Table 8 shows the numerical data for this test.

© ISO/IEC 2007 – All rights reserved

Table 8 – Multi-channel test 1, mean values and the lower and upper 95% confidence interval limits for the overall results per codec.

Codec ID	Mean	Lower	Upper
HE-AAC_MPS_A	90.810	89.760	91.861
HE-AAC_MC_A	88.754	87.610	89.897
HE-AAC_MPS_B	76.415	74.562	78.269
HE-AAC_MC_B	63.426	61.150	65.701
L2_MPS	92.146	91.091	93.202
L2_DPL2	57.638	55.663	59.614
REF	99.462	99.296	99.627
LP35	19.556	18.877	20.236

The following statements can be made concerning performance for all excerpts:

- MPEG Surround applied to an HE-AAC stereo core-coder at a total bitrate of 160kbps (HE-AAC_MPS_A) has a performance at the 95% level of significance equal to that of a discrete multi-channel coding by HE-AAC (HE-AAC_MC_A) at the same bitrate. Both systems provide a quality in the "Excellent" range on the MUSHRA scale.
- MPEG Surround applied to an HE-AAC stereo core-coder at a total bitrate of 64kbps (HE-AAC_MPS_B) performs better at the 95% level of significance than a discrete multi-channel coding by HE-AAC (HE-AAC_MC_B) at the same bitrate. For this operating point, MPEG Surround provides a very large advantage of more than 10 points on the MUSHRA scale.
- MPEG Surround applied to an MPEG-1 Layer II stereo core-coder at a total bitrate of 256kbps (L2_MPS) performs better at the 95% level of significance than a Dolby Prologic II encoding with MPEG-1 stereo core coder at a total bitrate of 256kbps (L2_DPL2). For this operating point, MPEG Surround provides a very large advantage of more than 25 points on the MUSHRA scale.

The plot in Figure 6 shows the results of the downmix test in which data from all test sites is pooled.

© ISO/IEC 2007 – All rights reserved

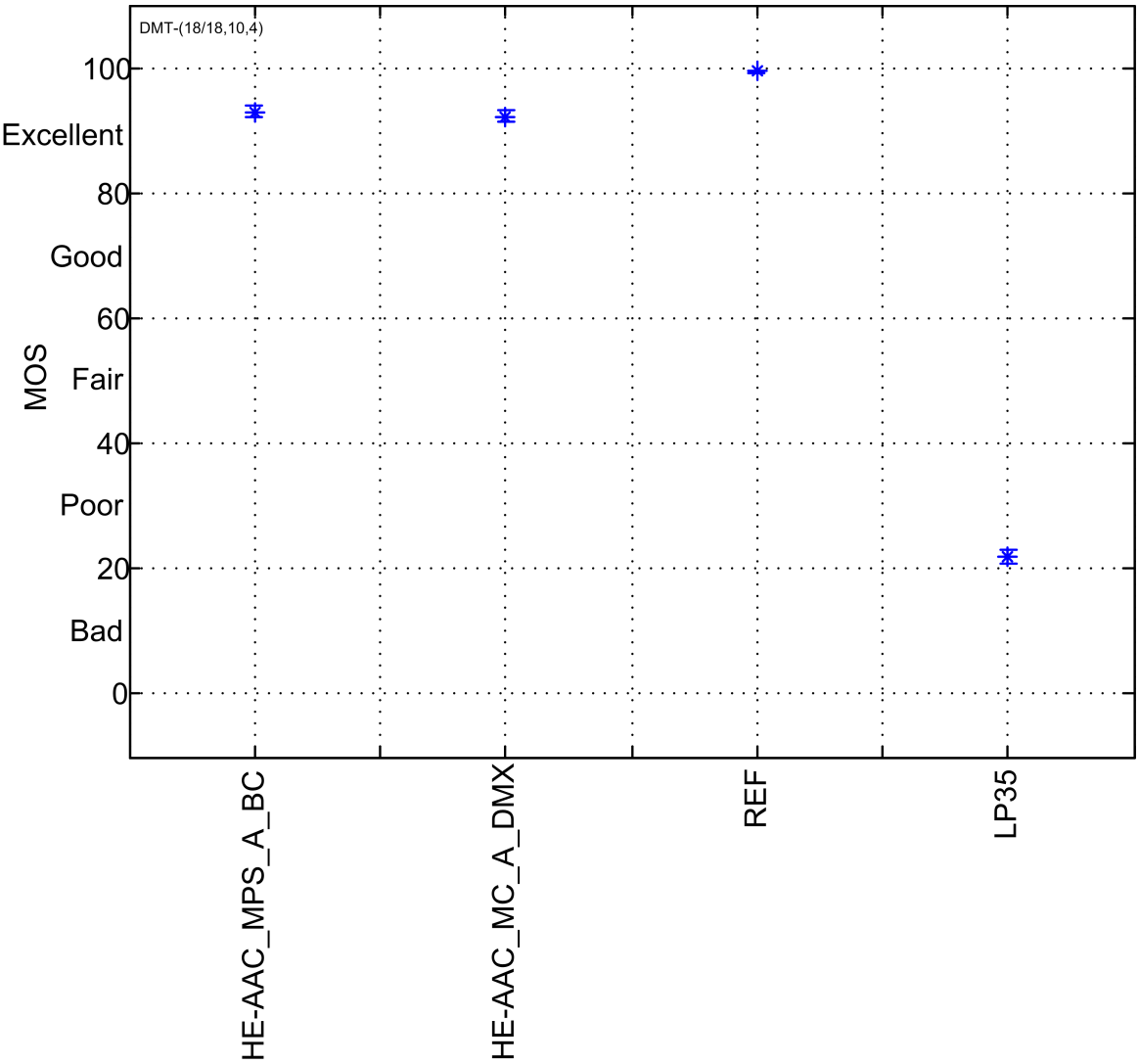

Figure 6 – Results for downmix test

Table 9 shows the numerical data for this test.

Table 9 – Downmix test , mean values and the lower and upper 95% confidence interval limits for the overall results per codec.

Codec ID	Mean	Lower	Upper
HE-AAC_MPS_A_BC	92.967	92.011	93.923
HE-AAC_MC_A_DMX	92.217	91.258	93.175
REF	99.306	99.005	99.606
LP35	21.939	20.937	22.940

The following statements can be made concerning performance for all excerpts:

- The inherent stereo downmix of the MPEG Surround system with HE-AAC as stereo core coder has equal quality equal to that of an ITU downmix of the HE-AAC discrete multi-channel coded output.

As outlined in Appendix 0, the HE-AAC stereo core coder in combination with MPEG Surround was operated at 83kbps. From the results it can be observed that when creating an automated downmix by means of the

© ISO/IEC 2007 – All rights reserved

MPEG Surround encoder, and coding this with 83kbps, the quality is equal to that obtained by coding discrete multi-channel at 160kbps, and subsequently performing the downmix.

In summary it can be concluded that excellent surround sound quality can be achieved using MPEG Surround in combination with both HE-AAC and MPEG-1 Layer 2. Observing the test results for HE-AAC, it can be concluded that for establishing a new high quality service, the provider can use the MPEG Surround to achieve the same quality as the discrete multi-channel coding using HE-AAC. By doing so, the service provider can easily serve both stereo and multi-channel receivers since the stereo signal is inherently part of the MPEG Surround codec signal. For the scenario where there are bitrate constraints (the 64 kbps scenario), a service provider can by using MPEG Surround, offer higher quality than available using discrete multi-channel coding. Furthermore, for upgrading a Layer-2 based system the only options previously available were matrix technology such as Dolby Prologic II, or simulcast at a high bitrate penalty. MPEG Surround in combination with Layer-2 offers a vastly superior alternative.

5.3 Music-store / portable player oriented

As outlined above, the music-store / portable player oriented test-case comprises multi-channel test 2 and a binaural test. The plot in Figure 7 shows the results of the subjective test of multi-channel test 2 in which data from all test sites is pooled.

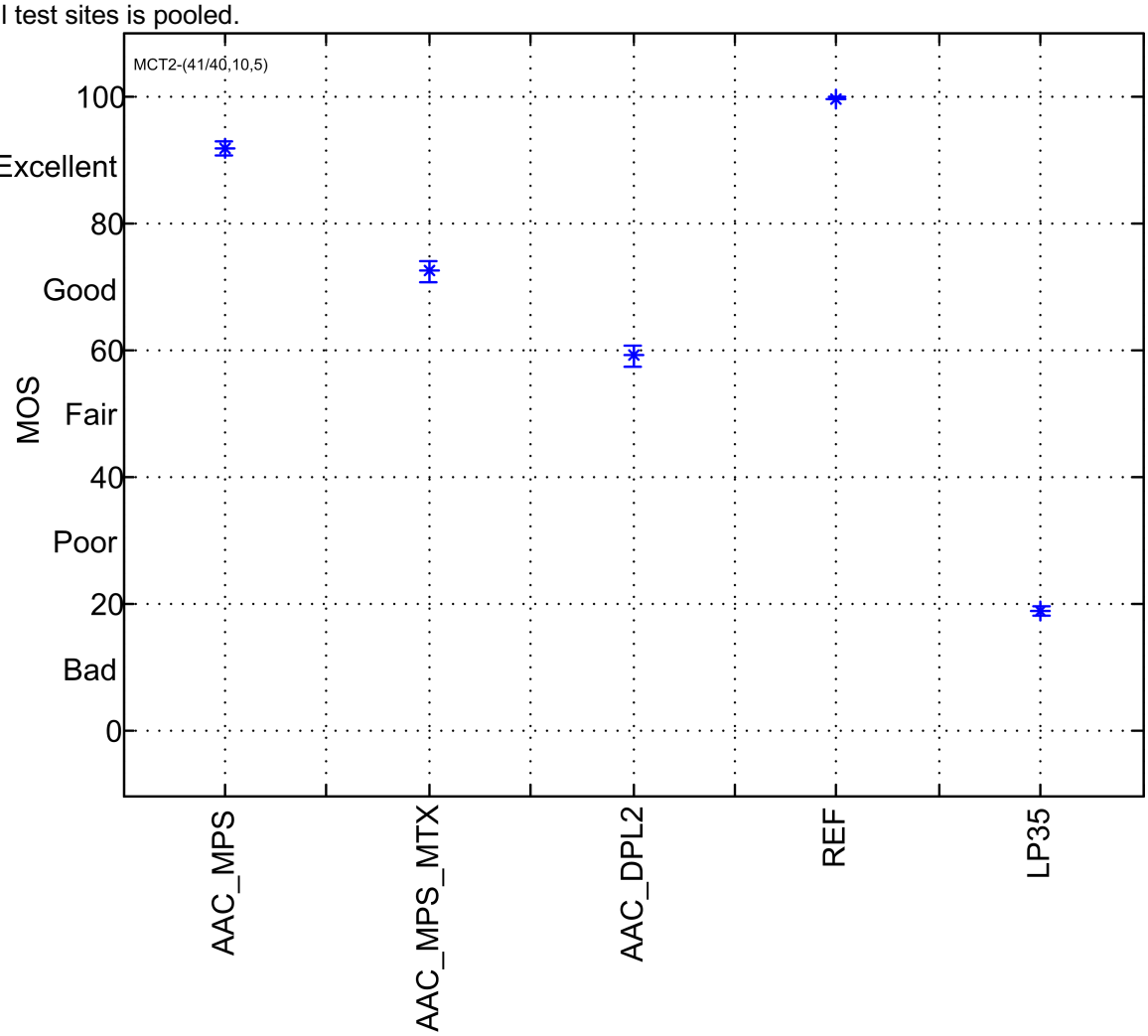

Figure 7 – Results for multi-channel test 2

Table 10 shows the numerical data for this test.

© ISO/IEC 2007 – All rights reserved

Table 10 – Multi-channel test 2, mean values and the lower and upper 95% confidence interval limits for the overall results per codec.

Codec ID	Mean	Lower	Upper
AAC_MPS	91.740	90.541	92.939
AAC_MPS_MTX	72.438	70.793	74.082
AAC_DPL2	59.072	57.321	60.824
REF	99.530	99.371	99.689
LP35	19.000	18.382	19.618

The following statements can be made concerning performance for all excerpts:

- MPEG Surround operated with AAC as a core coder performs better than both the Enhanced Matrix Mode operation of MPEG Surround and AAC in combination with Dolby Prologic II matrix technology.
- The Enhanced Matrix Mode of MPEG Surround performs better than the Dolby Prologic II matrix technology.
- MPEG Surround applied to an AAC stereo core-coder at a bitrate of 160 kbps (AAC_MPS) provides a quality in the "Excellent" range on the MUSHRA scale.

In summary it can be concluded that excellent surround sound quality can be achieved using MPEG Surround in combination with AAC. Furthermore, when en/decoding content in a matrix compatible way, the MPEG Surround enhanced matrix mode offers a decoding mode that provides quality superior to that of regular matrix decoding (even without access to side information).

The plot in Figure 8 shows the results of the subjective test of the binaural test in which data from all test sites is pooled.

© ISO/IEC 2007 – All rights reserved

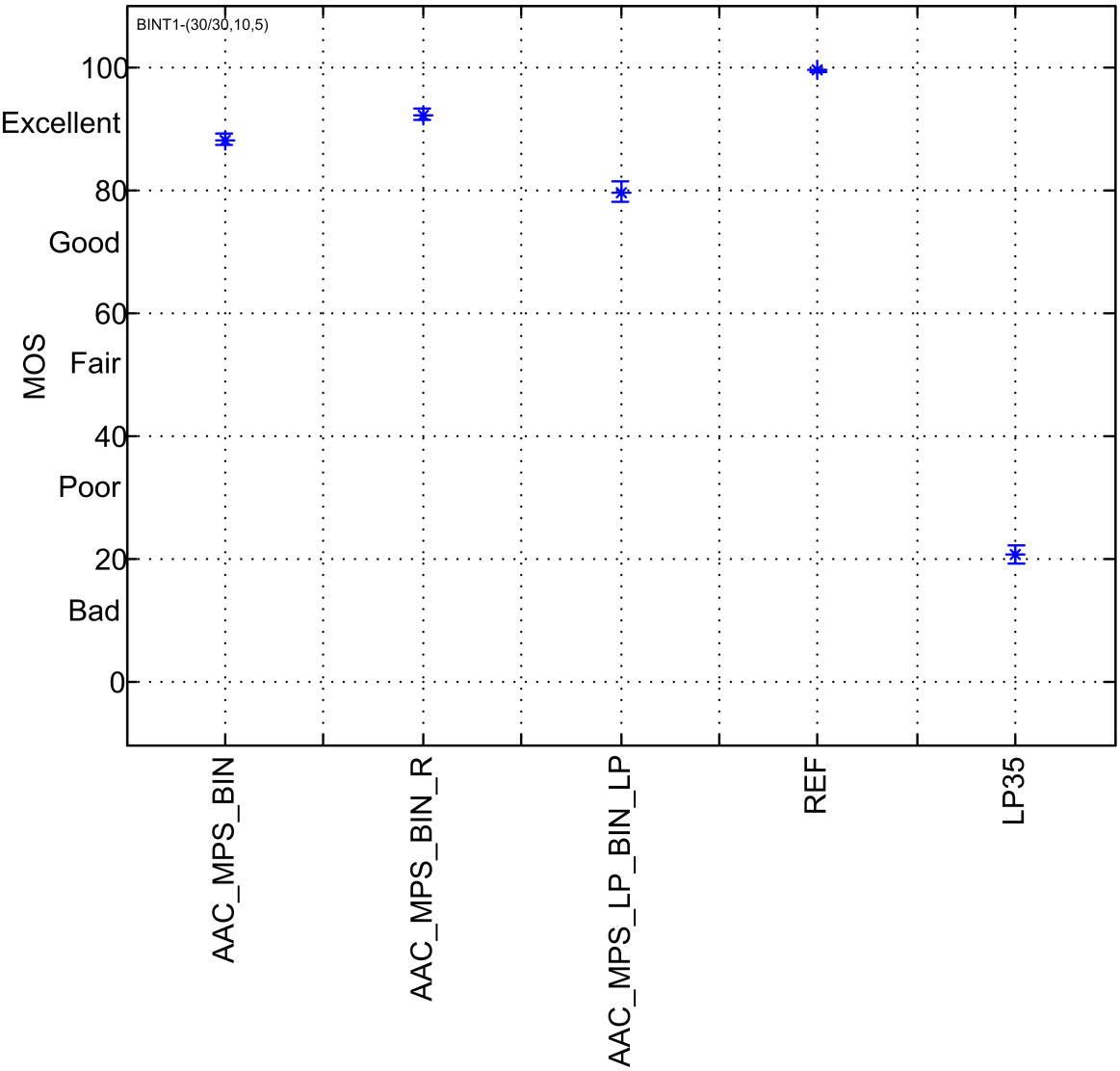

Figure 8 – Results for Binaural

Table 11 shows the numerical data for this test.

Table 11 – Binaural test, mean values and the lower and upper 95% confidence interval limits for the overall results per codec.

Codec ID	Mean	Lower	Upper
AAC_MPS_BIN	88.180	87.121	89.239
AAC_MPS_BIN_R	92.267	91.278	93.255
AAC_MPS_LP_BIN_LP	79.717	78.165	81.268
REF	99.353	99.137	99.569
LP35	20.750	19.335	22.165

The following statements can be made concerning performance for all excerpts:

© ISO/IEC 2007 – All rights reserved

- The binaural downmix application after MPEG Surround upmix performs better than the application of the binauralization directly on the stereo downmix in the binaural decoder.
- The high quality filtering approach of MPEG Surround binaural decoding performs better than the low power parametric MPEG Surround binaural decoding.

Both the high quality as well as the low complexity MPEG surround binaural reconstructions do not exploit the residual bitstream component that is present in MPEG Surround bitstream AAC_MPS. This can be accounted to the complexity/quality trade-offs made in MPEG Surround. However, in generating the binaural reference signal (AAC_MPS_BIN_R), this residual component has actually been used. This leads to a higher quality multi-channel reconstruction prior to the HRTF filtering as opposed to that available to the MPEG Surround binaural rendering engine. This explains the small difference in quality between the binaural reference and the high quality MPEG Surround binaural output.
The difference between the high quality and the low complexity MPEG Surround binaural output can be attributed to the fact that AAC_MPS_LP_BIN_LP employs a single reverberation (as imposed by the parallel reverb module) for all virtual channels while AAC_MPS_BIN employs different reverberation for each virtual channel. For further details, refer to Appendix A.4.2.

In summary, MPEG Surround in combination with the AAC core coder operated at a total bitrate of 160kbps offers a multi-channel audio quality in the "Excellent" range of the MUSHRA scale. The enhanced matrix mode of MPEG Surround operating without any MPEG Surround bitstream performs significantly better than Dolby Prologic II in combination with AAC. Finally, MPEG Surround offers several binaural rendering alternatives targeting different complexity all providing a binaural rendering quality in the "Excellent" range of the MUSHRA scale.

5.4 Technology driven test

As outlined in Section 2.1 a technology driven test was performed in order to assess the performance difference between the High Quality mode of operation versus the Low Power mode of operation of MPEG Surround. The technology driven tests include the multi-channel test 3 outlined in the figure below. Figure 9 plots the pooled data from all test sites.

© ISO/IEC 2007 – All rights reserved

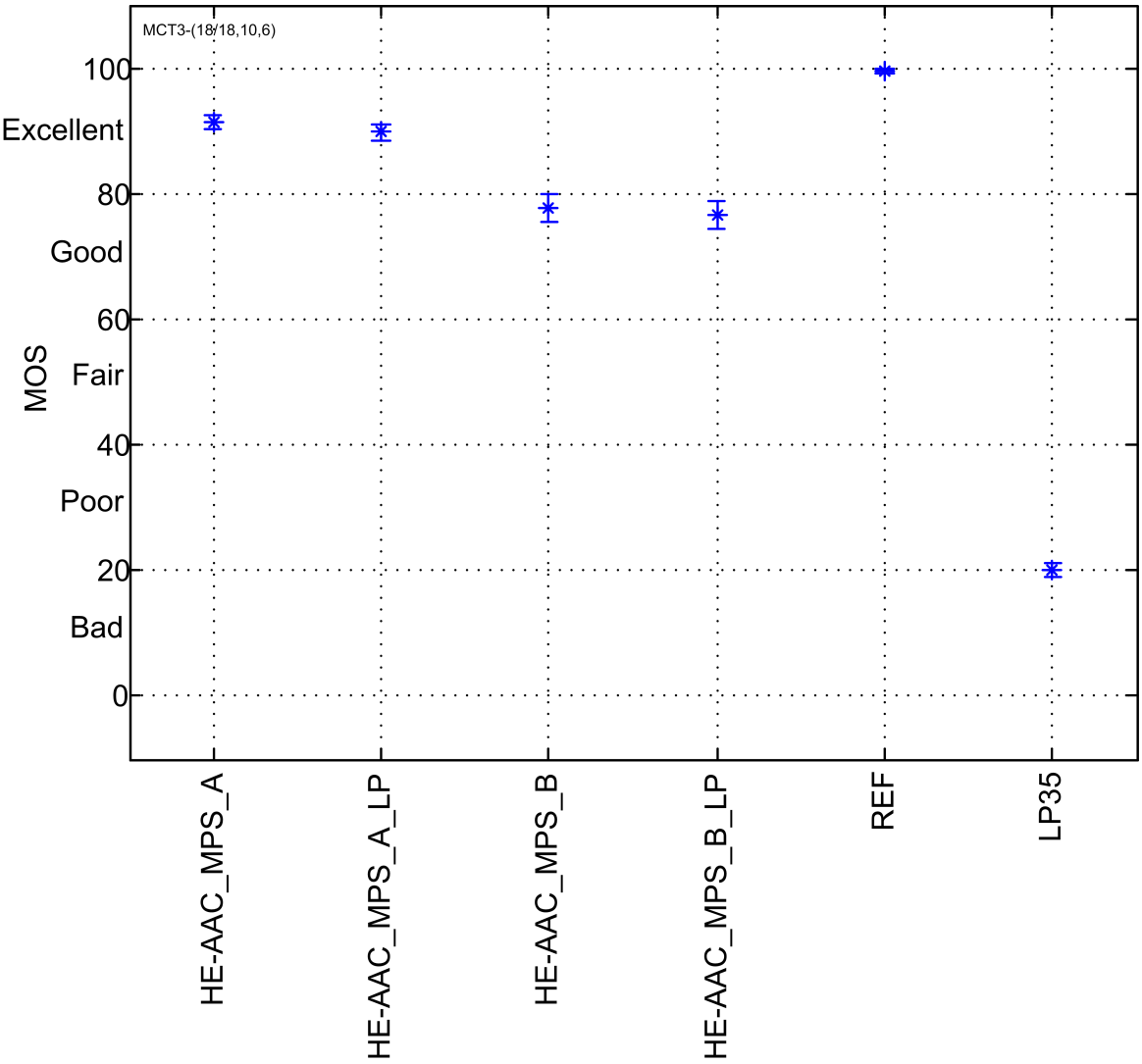

Figure 9 – Results for multi-channel test 3

Table 12 shows the numerical data for this test.

Table 12 – Multi-channel test 3, mean values and the lower and upper 95% confidence interval limits for the overall results per codec.

Codec ID	Mean	Lower	Upper
HE-AAC_MPS_A	91.372	90.113	92.632
HE-AAC_MPS_A_LP	89.822	88.478	91.166
HE-AAC_MPS_B	77.583	75.443	79.723
HE-AAC_MPS_B_LP	76.644	74.388	78.901
REF	99.478	99.209	99.746
LP35	20.056	19.061	21.050

The following statements can be made concerning performance for all excerpts:

© ISO/IEC 2007 – All rights reserved

For the test configurations used in the technology driven test no difference between the High Quality version and the Low Power version of MPEG Surround can be observed.

6 Conclusions

The verification test results shows that the MPEG Surround technology offers significant compression efficiency compared to discrete multi-channel coding at lower bitrates, and that it offers the same quality as discrete multi-channel coding at higher bitrates (in the "Excellent" range of the MUSHRA scale) while providing an excellent quality, inherently backwards compatible stereo downmix.

Furthermore, the MPEG Surround technology offers several modes of binaural rendering all providing excellent quality and enabling surround sound on portable devices.

Finally, the enhanced matrix mode of MPEG Surround operating without additional side-information performs vastly superior to the Dolby Prologic II technology.

The two application driven tests show that MPEG Surround is an ideal technology for broadcasting services where inherent stereo compatibility is vital. In the Music store / portable player scenario, MPEG Surround can deliver an excellent surround sound experience both at home on a multi-channel speaker set-up and on a portable device over headphones.

7 Acknowledgements

The following persons are acknowledged for their contribution to the verification tests:

Table 13 – Acknowledgements.

Name	Company
Schuyler Quackenbush	Audio Research Labs
Pierrick Philippe	France Télécom R&D
Johannes Hilpert	FhG-IIS
Toshiyuki Nomura	NEC
Naoya Tanaka	MEI
Chong Kok Seng	Panasonic
Jeongil Seo	ETRI
Hyunkook Lee	LG
Kristofer Kjörling	Coding Technologies
Oliver Niemeyer	Thomson Hannover
Jeroen Koppens	Philips Applied Technologies

8 References

[1] Multi stimulus test with hidden reference and anchor (MUSHRA) - EBU method for subjective listening tests of intermediate audio quality, Recommendation ITU-R BS. 1534 available at http://ecs.itu.ch

[2] Audio Subgroup **Call for Proposals on Spatial Audio Coding** ISO/IEC JTC1/SC29/WG11/N6455. March 2004

[3] Audio Subgroup **DRAFT Procedures for the Evaluation of Spatial Audio Coding Systems** ISO/IEC JTC1/SC29/WG11/N6456. March 2004

[4] Audio Subgroup **Procedures for the Evaluation of Spatial Audio Coding Systems** ISO/IEC JTC1/SC29/WG11/N6691. July 2004

© ISO/IEC 2007 – All rights reserved

[5] Audio Subgroup **Report on Spatial Audio Coding RM0 Selection Tests** ISO/IEC JTC1/SC29/WG11/N6813. October 2004

[6] Audio Subgroup **Workplan for MPEG Surround verification test** ISO/IEC JTC1/SC29/WG11/N8637. October 2006

[7] Audio Subgroup **Performance of MPEG Surround Technology** ISO/IEC JTC1/SC29/WG11/N7950. January 2006

© ISO/IEC 2007 – All rights reserved

A.1 Testing Schedule for the verification test

In preparation for the verification test the schedule given in Table 14 was agreed upon.

Table 14 – Time line for subjective testing.

What	When	Who
Setup ftp site for item submission and publicise access details	August 1, 2006	France Télécom R&D
Close of submission of pre-selection items and HRTF set	September 15, 2006	All
Provide coded items for pre-selection	September 30, 2006	CT/FhG-IIS/Philips
Report pre-selection of items	Input 78th MPEG meeting	France Télécom R&D
Finalize Verification test plan	78th MPEG meeting	Audio subgroup
Final item levelling	Friday November 10, 2006	France Télécom R&D
Provide bitstreams, decoder and decoded wave files, step scripts, delay alignment	Friday December 1, 2006	CT/FhG-IIS/Philips
Provide Excel format for reporting test results	Friday December 1, 2006	Audio Chair
Report subjective test results	Friday December 22, 2006	Test sites
Analysis and test report	Input 79th MPEG meeting Monday January 8, 2006	Audio Chair
Final verification test report	Output 79th MPEG meeting Friday January 19, 2007	Audio subgroup

© ISO/IEC 2007 – All rights reserved

A.2 Pre-selection process

A.2.1 Collection process

Prior to the 78th MPEG meeting, a total of 159 items have been collected by France Télécom R&D as a base for further selection. The collected items have been encoded with HE-AAC+MPEG Surround at 64kbps and HE-AAC discrete at 64kbps. The encoding was done by Coding Technologies.

A.2.2 Pre-selection process

A selection panel at France Télécom R&D) has selected 10 verification test items + 3 items for training purposes that are critical for the systems under test. Prior to the actual selection, items have been categorized according to different genres (Movie/Drama, Orchestra, Pop, Ambiance, Jazz).

The main artefacts have been identified over these genres for the coders used for the pre-selection. Subsequently, the various artefacts have been balanced over the different genres when selecting items in order to prevent that a single artefact appears in every selected items (e.g. applause artefact).

Figure 10 illustrates the selected items (marked as red points) among the complete pre selection set. HE-AAC+MPS and HE-AAC multi-channel items were rated independently. The selected items were picked to be representative of the spreading of the scores obtained for the two codecs. This has been done for each genre separately.

Additionally, the following criteria have been applied in order to arrive at the selected items.

- items used during the core experiment process were removed
- Duplicate signals or recording were avoided
- Synthetic mixtures that do not reflect the targeted test scenarios were eliminated
- Less critical items were removed; similarly pathological items were avoided as well.

© ISO/IEC 2007 – All rights reserved

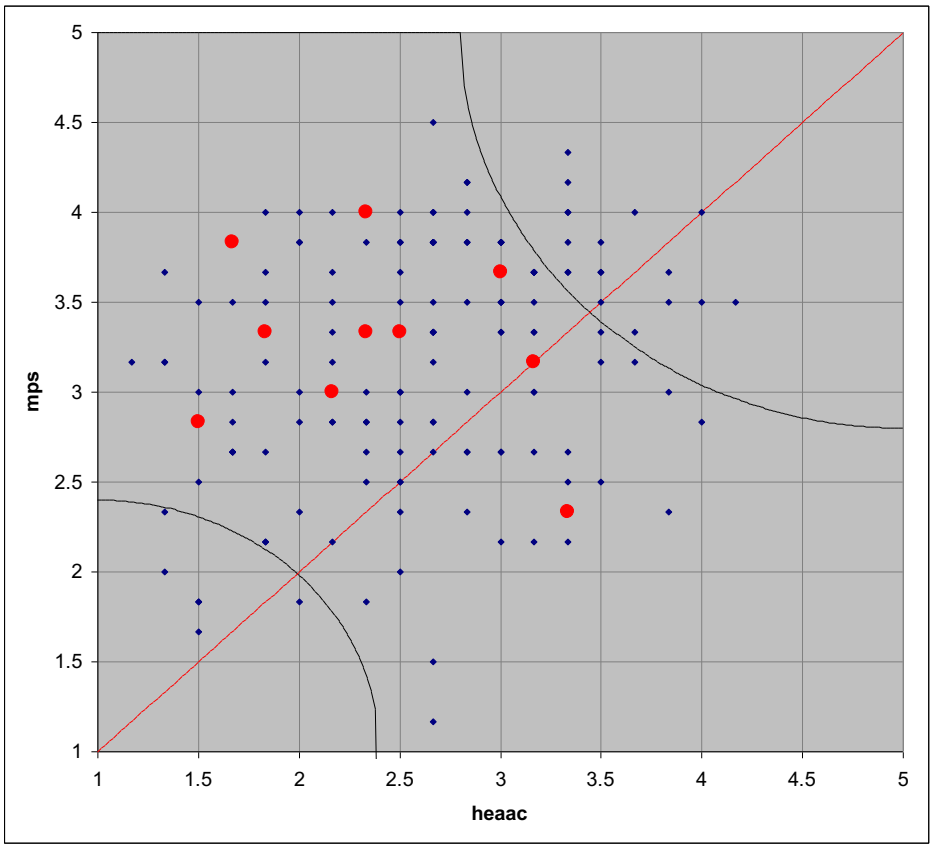

Figure 10 – Item pre-selection.

A.2.3 Test items

Table 15 lists the items that have been selected for the different tests. All Items are in 5.1 format and 44.1 kHz sampling rate and approximately 20-25 seconds in duration.

Table 15 – Items to be used for each test.

Item no.	Item name	Genre	Provided by	Duration (s)	LFE
11	Station_Atmo_6ch	Ambience	Uni. Dusseldorf	20	
15	tennis	Ambience	NBC	24	
32	tower1	Jazz	NBC	17	
36	bonobo_1	Movies Drama	France Telecom	16	
40	elliot2	Movies Drama	NBC	21	
64	ravel1c	Orchestra	NBC	19	
77	violin2	Orchestra	NBC	21	
96	Lavilliers_14	Pop (Lavilliers)	France Telecom	22	x
134	thalheim4	Pop	NBC	19	
114	SantaCruz_09a	Pop (SantaCruz)	France Telecom	19	

Table 16 lists the items that are to be used for the training.

© ISO/IEC 2007 – All rights reserved

Table 16 – Training items

Item no.	Item name	Genre	Provided by	Duration (s)	LFE
16	chostakovitch	Core Experiment	NBC	26	
17	fountain_music	Core Experiment	NBC	21	
20	jackson1	Core Experiment	NBC	16	

A set of HRTFs (JBears), has been provided by Philips. This set has been used in several demonstrations and has shown to provide a clear virtual impression for a large number of people. This specific HRTF is available in the MPEG Surround reference decoder as one of the HRTFs that can be selected. For more information on the derivation and application of the JBears set of HRTFs, please refer to A.4.2.

© ISO/IEC 2007 – All rights reserved

A.3 Test methodology

A.3.1 MUSHRA

The tests in this report used the MUSHRA method described in [1]. This was developed in 1999 by EBU Project Group B/AIM, in collaboration with ITU-R Working Party 6Q. An important feature of this method is the inclusion of the hidden reference and two bandwidth limited anchor signals (7 kHz and 3.5 kHz).

A quality scale is used where the intervals are labeled "bad", "poor", "fair", good" and "excellent" as opposed to BS.1116. The value on the lower end of the scale is zero, the value on the upper end is 100. No decimals are given. This scale has the advantage to be harmonized with video quality.

The length of the sequences did not exceed 24 seconds to avoid fatiguing listeners and to reduce the total duration of the listening test.

A.3.2 Training phase

In order to obtain reliable results, it was mandatory to train the subjects in special training sessions in advance of the test. In preparation of the test, the subjects received both explanations and instructions about the test.

The purpose of the training phase was to allow the subject to achieve two objectives as follows:

- Become familiar with all the sound excerpts under test and their quality level ranges;
- Learn how to use the test equipment and the grading scale

During the training phase, the subject is able to listen to the 3 sound excerpts that have been selected in the course of the pre-selection process (see Table 16).

During the training phase, the subject was asked to use the available scoring equipment and evaluate the quality of the sound excerpts by inputting the appropriate scores on the continuous quality scale.

The subjects were instructed that they should not necessarily give grade "Bad" to the sound excerpt with lowest quality, or grade "Excellent" to the sound excerpt with highest quality with the exception of the hidden reference that has to be graded on top of the scale. This means, **at least one out of all test items had to be graded on top of the scale.** Besides this constraint they should use the range they find appropriate.

During the training phase the subjects were able to learn how they should interpret the audible impairments in terms of the grading scale. No grades given during the training phase were taken into account in the real tests.

The purpose of the grading phase was to input individual scores in the quality scale and to get used to the user interface. The scores should reflect the subjective judgment of the quality level for each of the sound excerpts presented. During the training phase, the subjects had to run through all the tested conditions.

The subject could discuss only the perceived artifact with the test administrator but not the specific grades in order to avoid bias in individual grading.

A.3.3 User - interface

Compared to ITU-R BS.1116, the MUSHRA method has the advantage of displaying all stimuli (conditions) for one test item. The subjects were therefore able to carry out any comparison between them directly.

Figure 11 shows the user-interface presenting one item under test. The buttons represent the reference, which is specially displayed on bottom left, and all the codecs under test, including the hidden reference and the anchor point (band-limited processed reference), called test items. Above each button, with the exception

© ISO/IEC 2007 – All rights reserved

of the button for the reference, a slider was used to grade the quality of the test item according to the continuous quality scale used. For each of the items, the signals under test were randomly assigned. In addition, the test items were randomized for each subject within a session. To avoid sequential effects, each subject was running the two sessions in randomized order.

None of the subjects had the same items order and the same order in the conditions presentation.

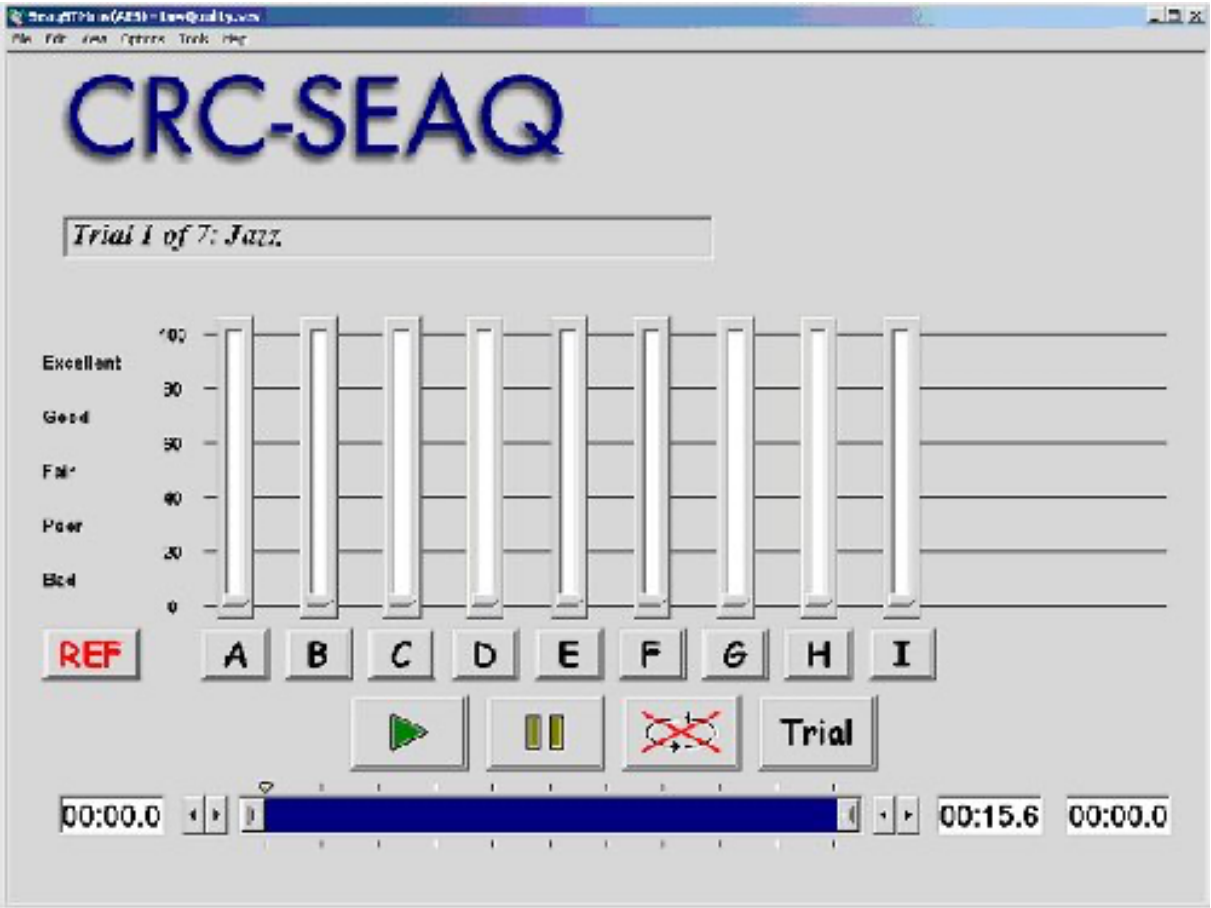

Figure 11 - User interface for MUSHRA tests

A.3.4 Listening time

Table 17 lists indicative listening time required for the 5 different tests outlined in this document.

Table 17 – Indicative timing for testing.

Test	#stimuli	#items	#of times listening	Duration (min)	Listening time for 10 subjects (hrs)
MCT1	8	10	4	107	18
DMT	4	10	2	27	4.5
MCT2	5	10	3	50	8.5
BINT	5	10	3	50	8.5
MCT3	6	10	3	60	10

© ISO/IEC 2007 – All rights reserved

A.4 Codec settings

A.4.1 Encoder settings

The following encoder settings have been used for coding the various test conditions.

Table 18 – Applied encoder settings for generating the test bitstreams.

Codec ID	Total bit-rate [kbps]	Core bit-rate [kbps]	Total spatial bit-rate [kbps]	Bitrate for residual [kbps]	Settings
HE-AAC_MPS_B	64	56	8	0	GES, Fractional Delay
HE-AAC_MPS_B_LP	64	56	8	0	GES
HE-AAC_MPS_A	160	83	77	3x25	GES, Fractional Delay
HE-AAC_MPS_A_LP	160	83	77	3x25	GES
AAC_MPS	191	160	31	1x10+2x5	Matrix mode
L2_MPS	256	225	31	1x10+2x5	GES, Fractional Delay

Dolby Prologic II encoding and decoding was performed by FhG-IIS. The Layer 2 encoding was done by Coding Technologies. The other encoding tasks were shared by Coding Technologies, FhG-IIS and Philips.

A.4.2 Binaural decoding settings

The "JBears" HRTFs have been measurement using in-ear microphones in an echoic room using a 5.0 ITU speaker setup. The speakers were placed symmetrically in a room without any obstacles. The HRTFs have been measured by alternating playing noise responses though loudspeakers and measuring 4096 taps impulse responses (a set of 2x5 responses). These HRTFs have been split into the direct and reverberant portion. Subsequently, spectral corrections have been applied in order to compensate for microphone characteristics to provide for more pleasant overall sounding HRTFs. Note that these were separate corrections for direct and early reflections. Finally the HRTFs have been cropped to 2048 samples.

For the high quality binaural mode the reverb component is different for every virtual speaker, as imposed by the employed set of HRTFs.

For the low power parametric implementation of the HRTFs, the echoic portion of the HRTF has been implemented in a parallel reverb module. This module inputs the L+R downmix signal and applies a single reverb, the outputs of which are added to the anechoic binaural signal. The parallel reverb module applies the reverb from the center channel HRTFs for all virtual channels (see Figure 12). The center channel reverb has been selected since it has a symmetrical position to both ears. Note that the parallel reverb is an informative part of the specification.

© ISO/IEC 2007 – All rights reserved

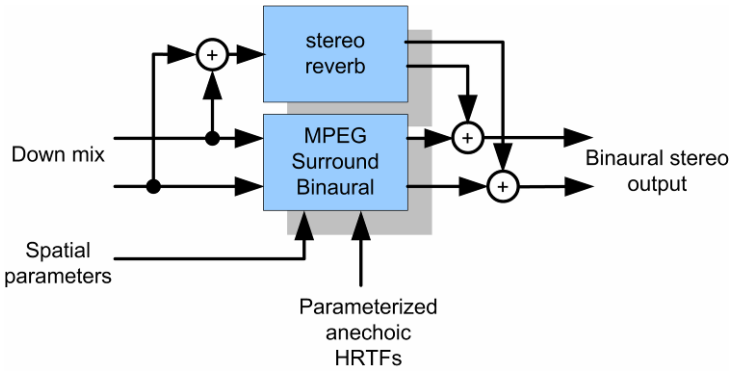

Figure 12 – Implementation of parallel reverb in MPEG surround.

© ISO/IEC 2007 – All rights reserved

A.5 Statistical Analysis

A.5.1 General

The statistical analysis followed standard MUSHRA procedure. The calculation of the averages of the scores of all listeners remaining after post-screening will result in the Mean Subjective Scores (MSS). The first step of the analysis of the results is the calculation of the mean score $\bar{u}_{jk}$, for each of the presentations:

$$\bar{u}_{jk} = \frac{1}{N}\sum_{i=1}^{N} u_{ijk}$$

where:

u_i is the score of observer i for a given test condition j and sequence k

N is the number of observers

Confidence intervals were also calculated which was derived from the standard deviation and the size of each sample. The 95% confidence interval is given by:

$$\left[\bar{u}_{jk} - \delta_{jk}, \bar{u}_{jk} + \delta_{jk}\right],$$

where:

$$\delta_{jk} = 1.96\frac{S_{jkl}}{\sqrt{N}},$$

and the standard deviation S_{jk} is given by: $S_{jk} = \sqrt{\sum_{i=1}^{N}\frac{\left(\bar{u}_{jk} - u_{ijk}\right)^2}{\left(N-1\right)}}$.

With a probability of 95%, the absolute value of the difference between the experimental mean score and the "true" mean score (for a very high number of observers) is smaller than the 95% confidence interval, on condition that the distribution of the individual scores meets certain requirements.

Similarly, a standard deviation could be calculated for each test condition. It is noted however that this standard deviation will, in cases where a small number of test sequences are used, be influenced more by differences between the test sequences used than by variations between the assessors participating in the assessment.

A.5.2 Post-screening

In order to assess the subjects reliability a post-screening procedure was applied. Subjects satisfying one or both of the following criteria have been removed from the corresponding test in which they participated.

- If the subjects scored a reference stimuli below a score of 90,
- If the subject scores the anchor above the hidden reference.

Applying these criteria leads to the following number of subjects for each of the tests (see Table 7). All the results in this document are presented with this post-screening rule applied.

© ISO/IEC 2007 – All rights reserved

Table 19 – Number of subjects as a result of post-screening.

Number of subjects	Subjects in test				
	Multi-channel test 1	Downmix test	Multi-channel test 2	Binaural test	Technology driven test
Without post-screening applied	41	18	41	30	18
With post-screening applied	39	18	40	29	18

© ISO/IEC 2007 – All rights reserved

A.6 Detailed Results all test sites

This section shows the results on a per item basis for all tests.

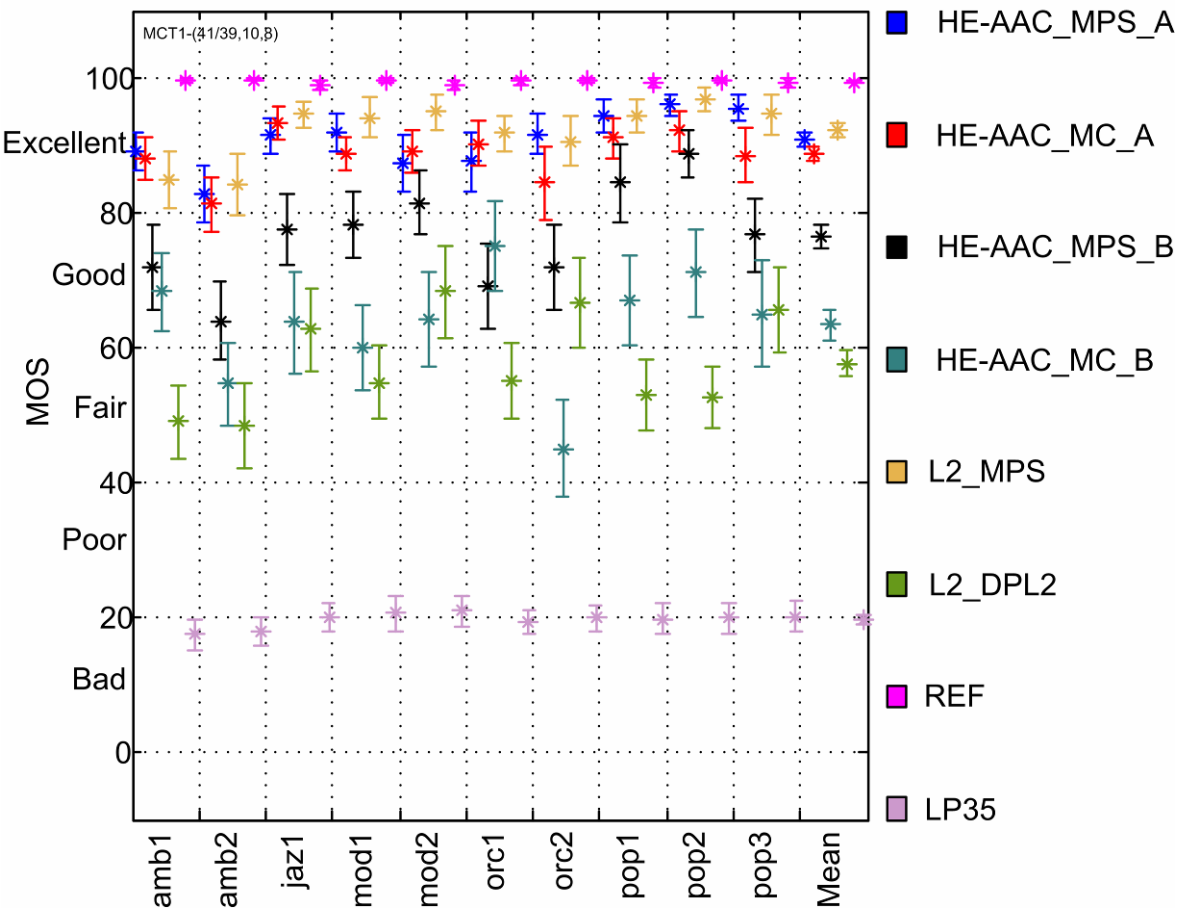

Figure 13 – Results for multi-channel test 1

© ISO/IEC 2007 – All rights reserved

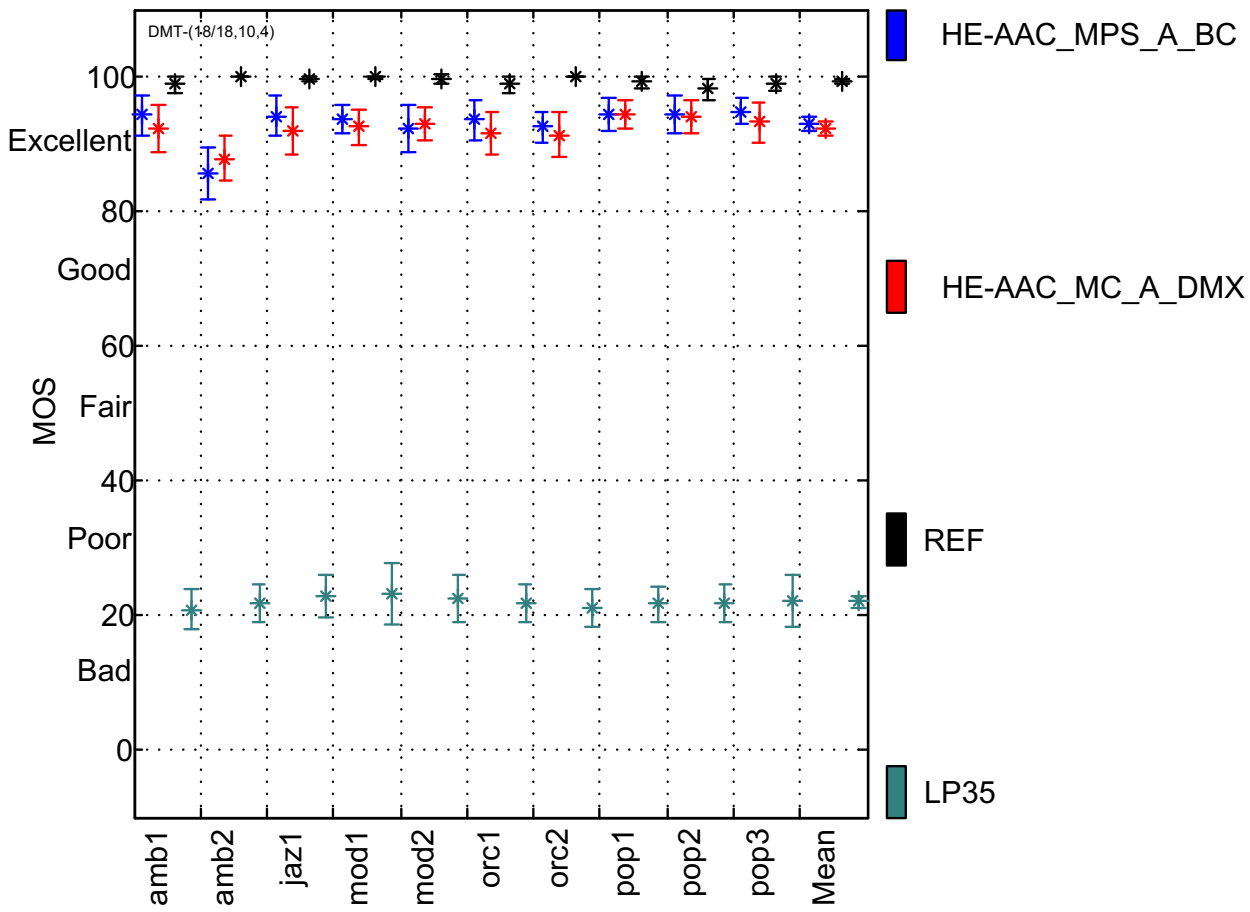

Figure 14 – Results for downmix test

© ISO/IEC 2007 – All rights reserved

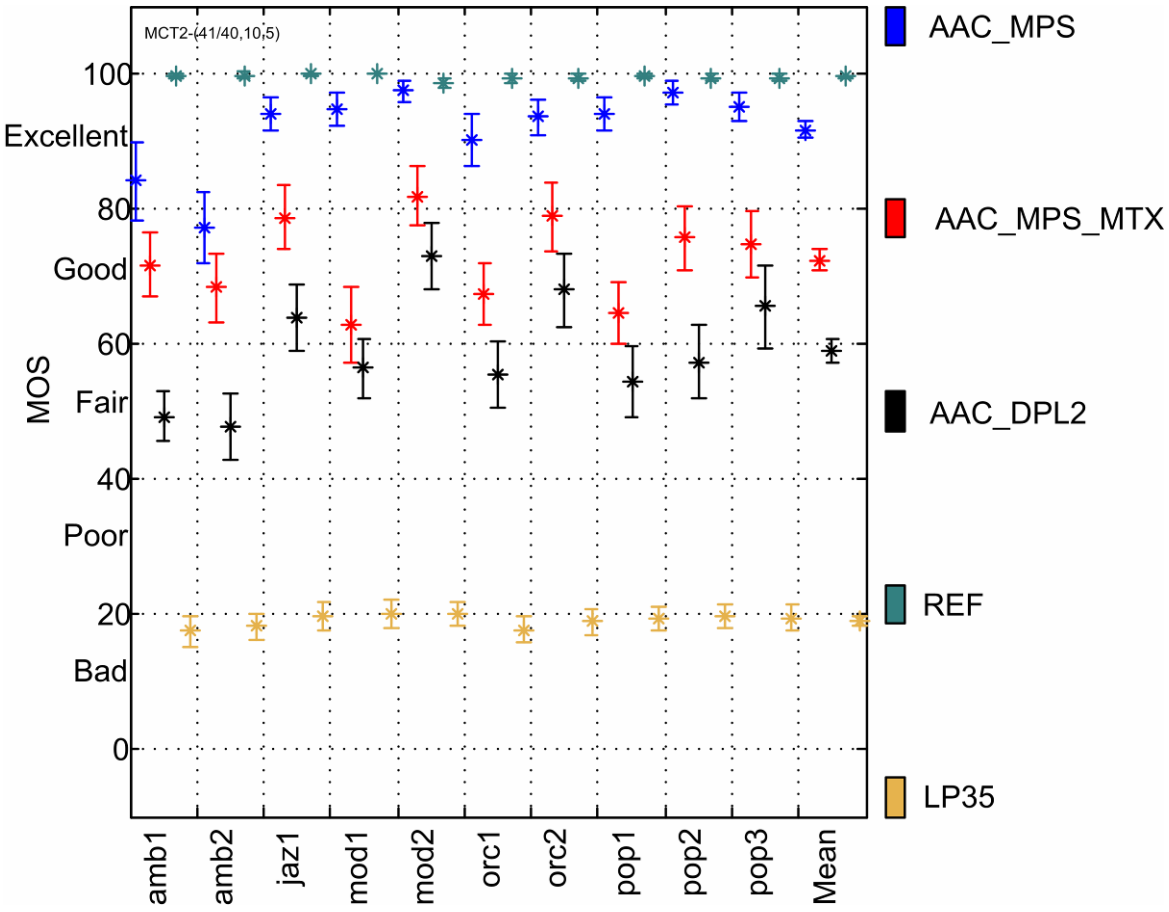

Figure 15 – Results for multi-channel test 2

© ISO/IEC 2007 – All rights reserved

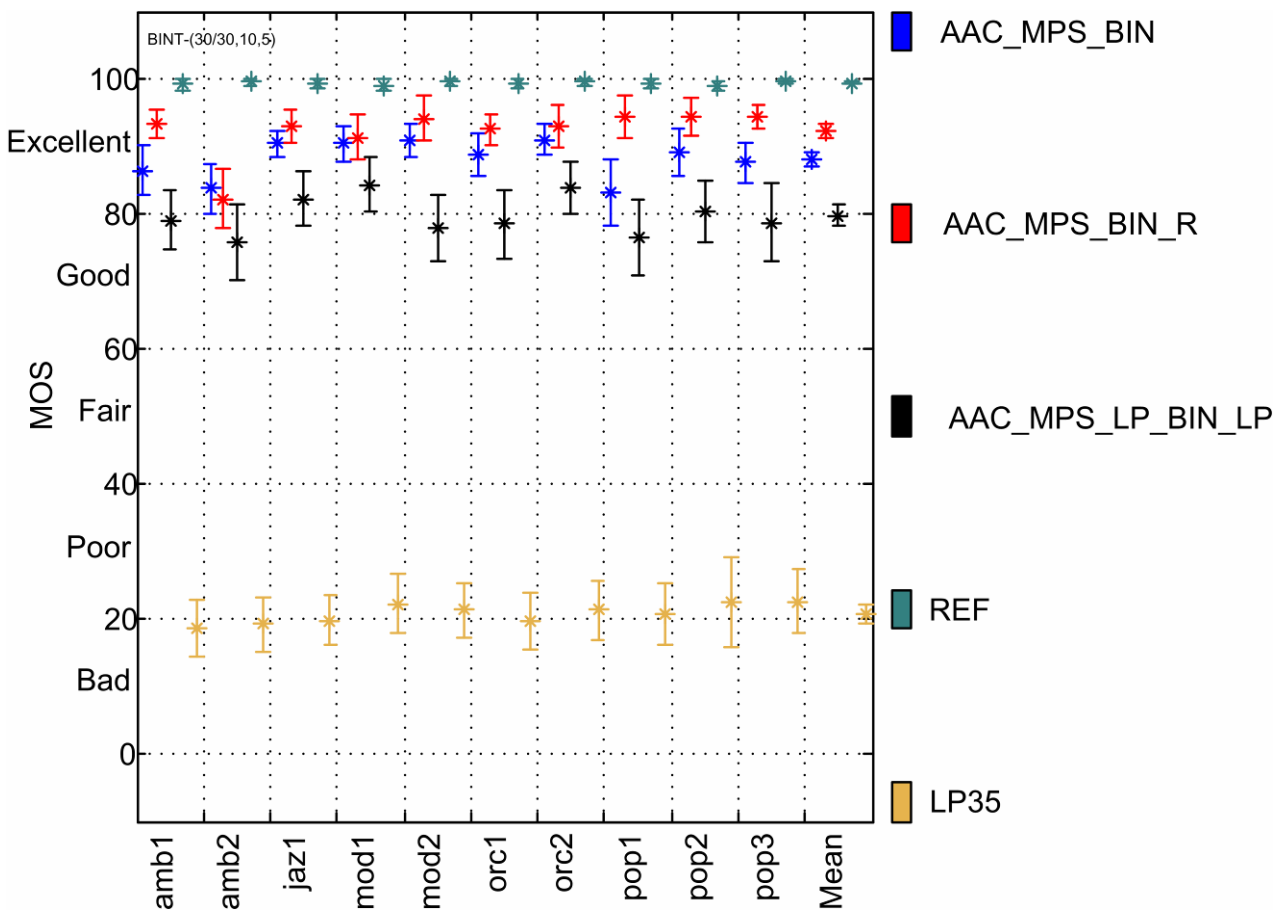

Figure 16 – Results for binaural test

© ISO/IEC 2007 – All rights reserved

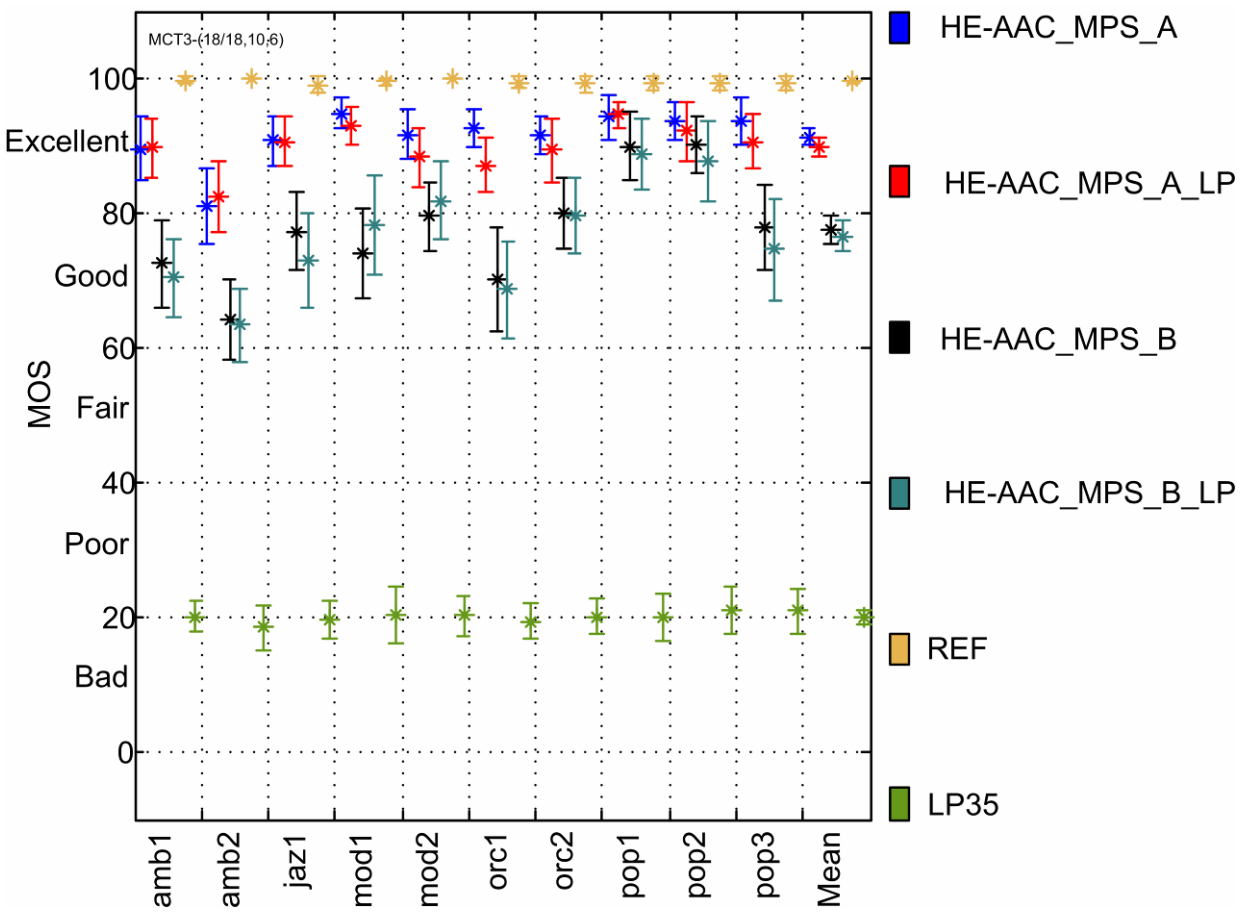

Figure 17 – Results for multi-channel test 3.

© ISO/IEC 2007 – All rights reserved

A.7 Detailed Results individual test sites

In order to assess the differences that occur over test sites, an ANOVA analysis has been performed for each test. Given that all p values for the various factors (i.e., test site, excerpt and codec) are below 0.007, the null hypothesis that these factors have no significant effect on the test results is rejected. In order to visualize this, for each test, Figure 18 shows the contrasts over the various test sites that participated in that test. Each data point represents the mean and 95% confidence interval for a specific test site over all items and codecs in a test.

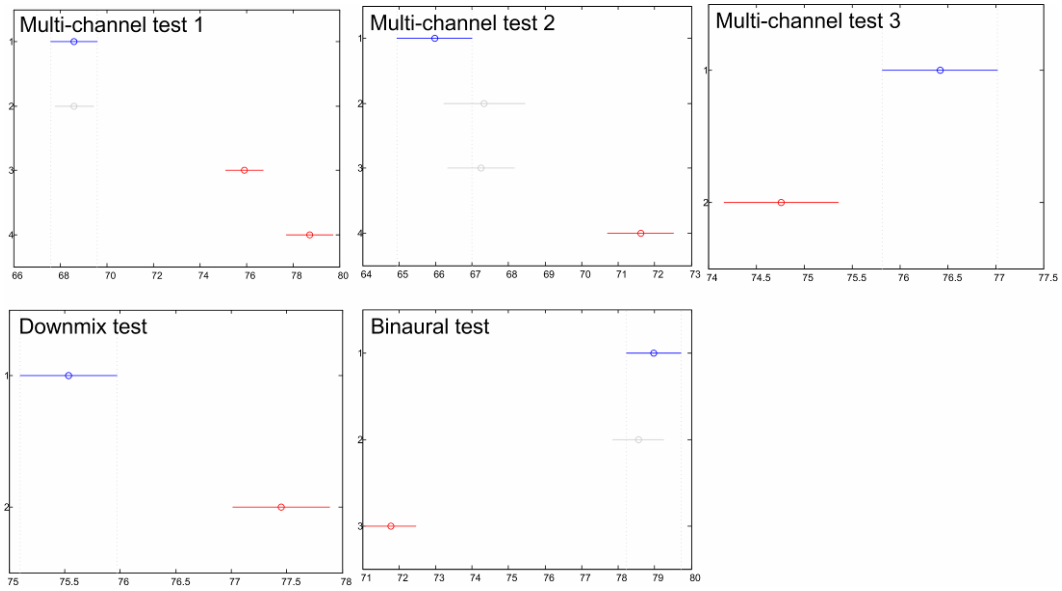

Figure 18 – ANOVA contrasts of test sites for all tests. Each line represents a test site.

A.7.1 Multi-channel test 1

This section provides both the overall results as well as the results on a per item basis per test site for multi-channel test 1. In addition also the mean and confidence intervals are tabulated.

© ISO/IEC 2007 – All rights reserved

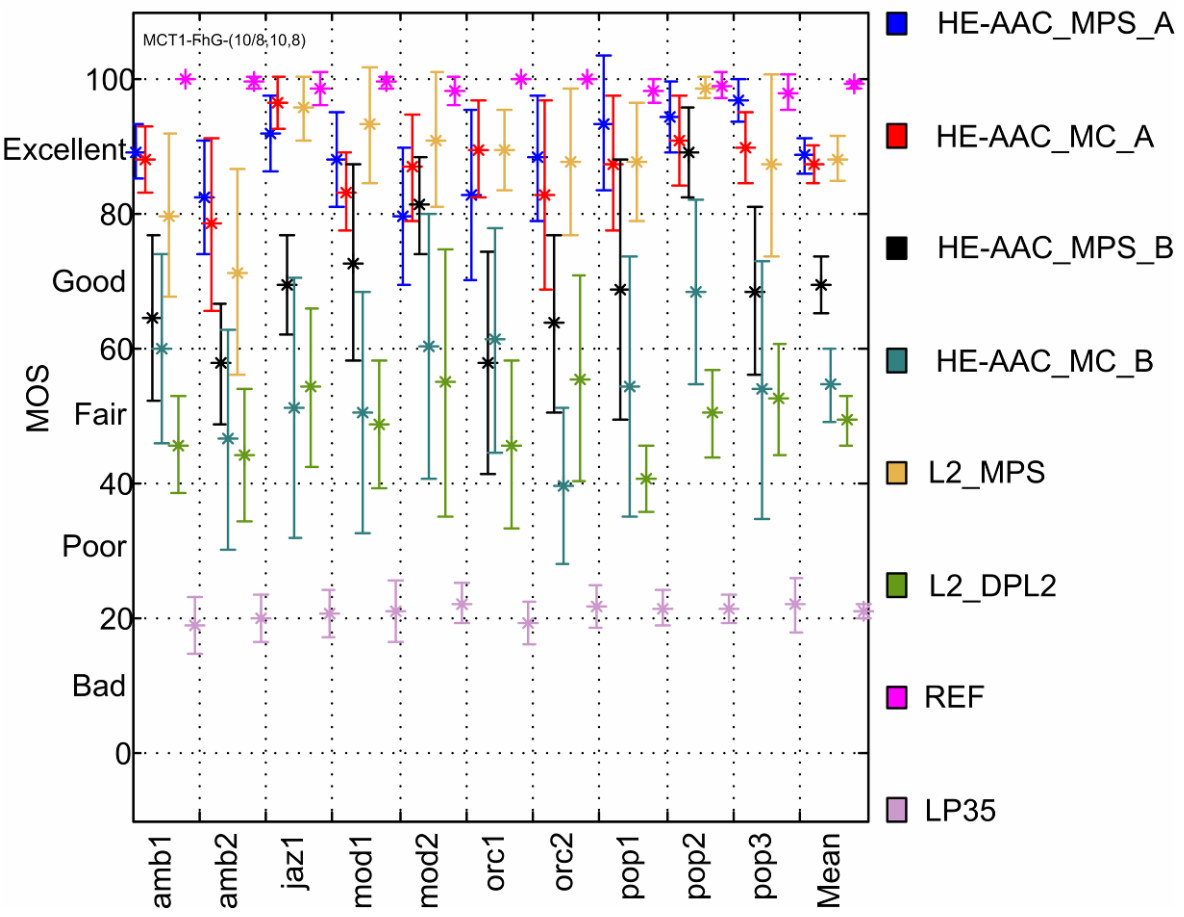

Figure 19 – Results for multi-channel test 1, test site FhG-IIS

© ISO/IEC 2007 – All rights reserved

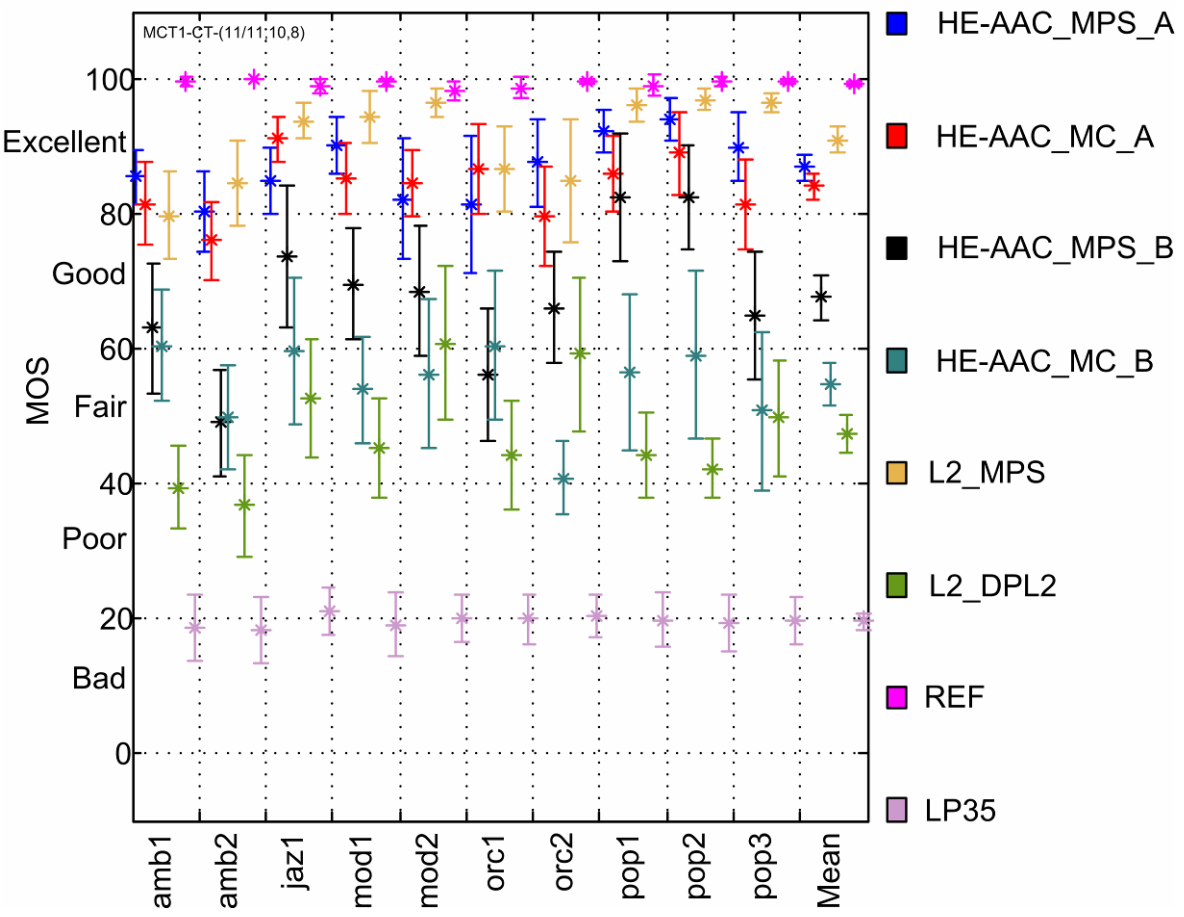

Figure 20 – Results for multi-channel test 1, test site CT

© ISO/IEC 2007 – All rights reserved

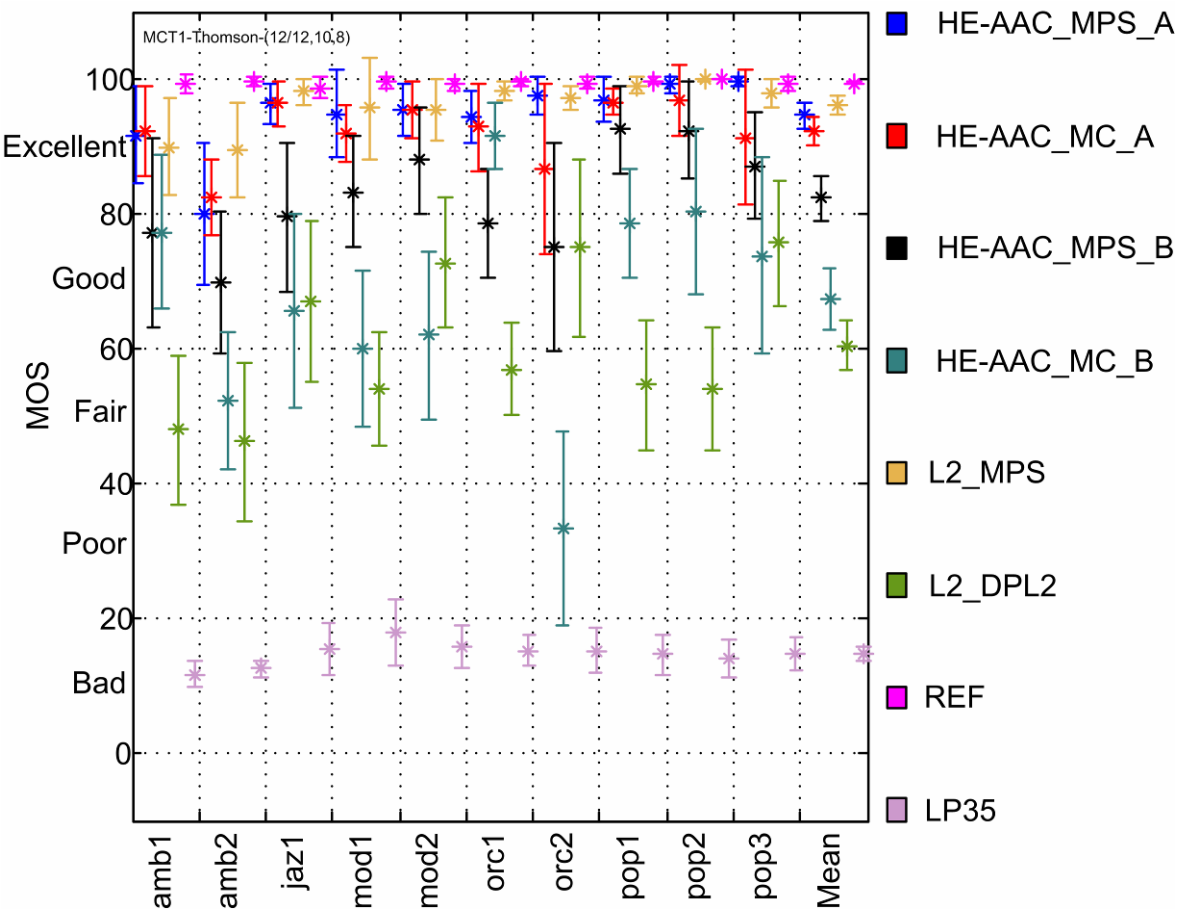

Figure 21 – Results for multi-channel test 1, test site Thomson Hannover

© ISO/IEC 2007 – All rights reserved

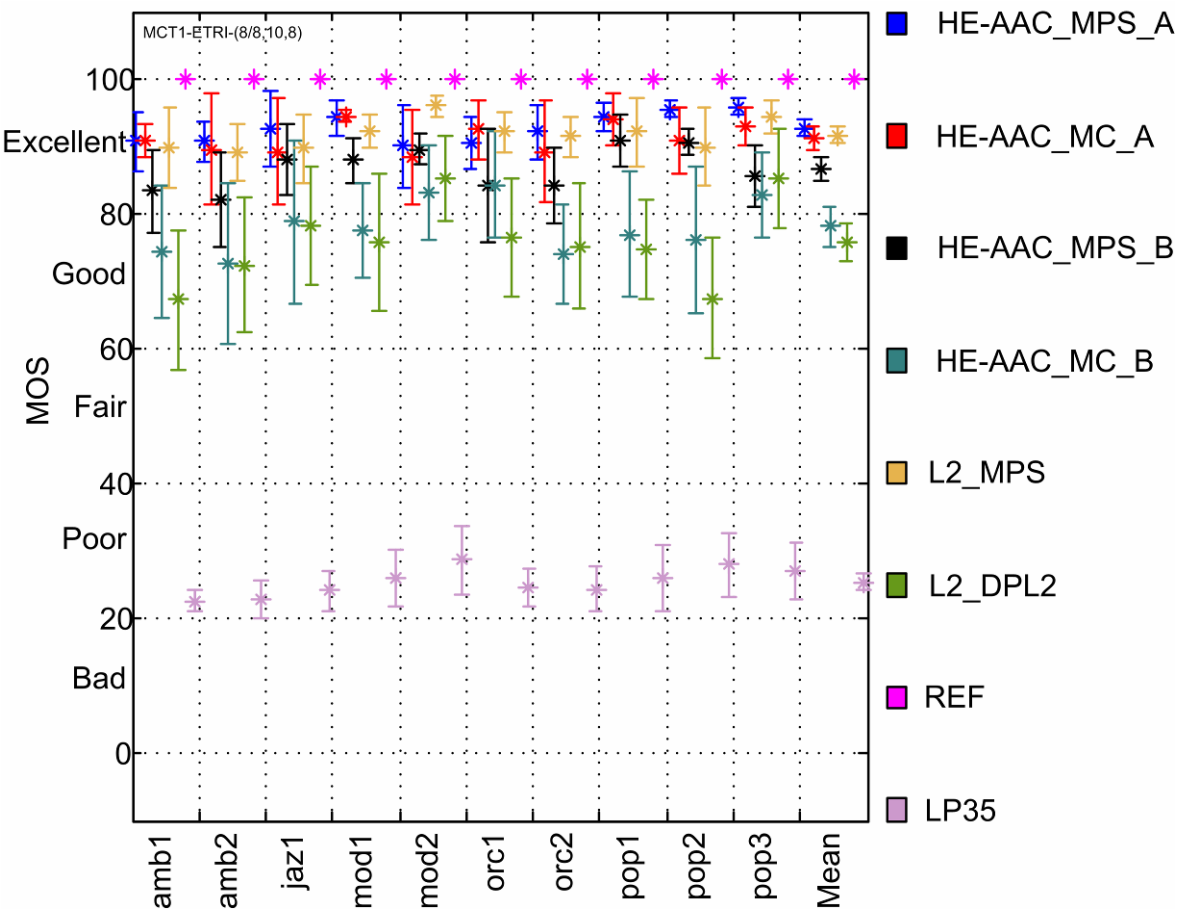

Figure 22 – Results for multi-channel test 1, test site ETRI

Table 20 –Overall results per test site.

Codec ID	Results per test site, Mean (lower,upper)			
	FhG-IIS	CT	Thomson Hannover	ETRI
HE-AAC_MPS_A	88.675 (85.983,91.367)	86.845 (84.825,88.866)	94.608 (92.781,96.436)	92.700 (91.493,93.907)
HE-AAC_MC_A	87.375 (84.636,90.114)	84.127 (82.170,86.085)	92.242 (90.041,94.442)	91.263 (89.540,92.985)
HE-AAC_MPS_B	69.375 (65.125,73.625)	67.600 (64.268,70.932)	82.358 (79.019,85.698)	86.662 (84.965,88.360)
HE-AAC_MC_B	54.625 (49.289,59.961)	54.736 (51.502,57.970)	67.492 (62.917,72.067)	78.075 (75.201,80.949)
L2_MPS	88.200 (84.818,91.582)	91.018 (89.140,92.896)	96.108 (94.639,97.578)	91.700 (90.379,93.021)
L2_DPL2	49.300 (45.767,52.833)	47.464 (44.580,50.347)	60.417 (56.773,64.060)	75.800 (72.877,78.723)
REF	99.125 (98.613,99.637)	99.355 (99.042,99.667)	99.425 (99.134,99.716)	100.000 (100.000,100.000)
LP35	20.900 (19.854,21.946)	19.609 (18.367,20.851)	14.725 (13.749,15.701)	25.387 (24.179,26.596)

A.7.2 Downmix test

This section provides both the overall results as well as the results on a per item basis per test site for the downmix test. In addition also the mean and confidence intervals are tabulated.

© ISO/IEC 2007 – All rights reserved

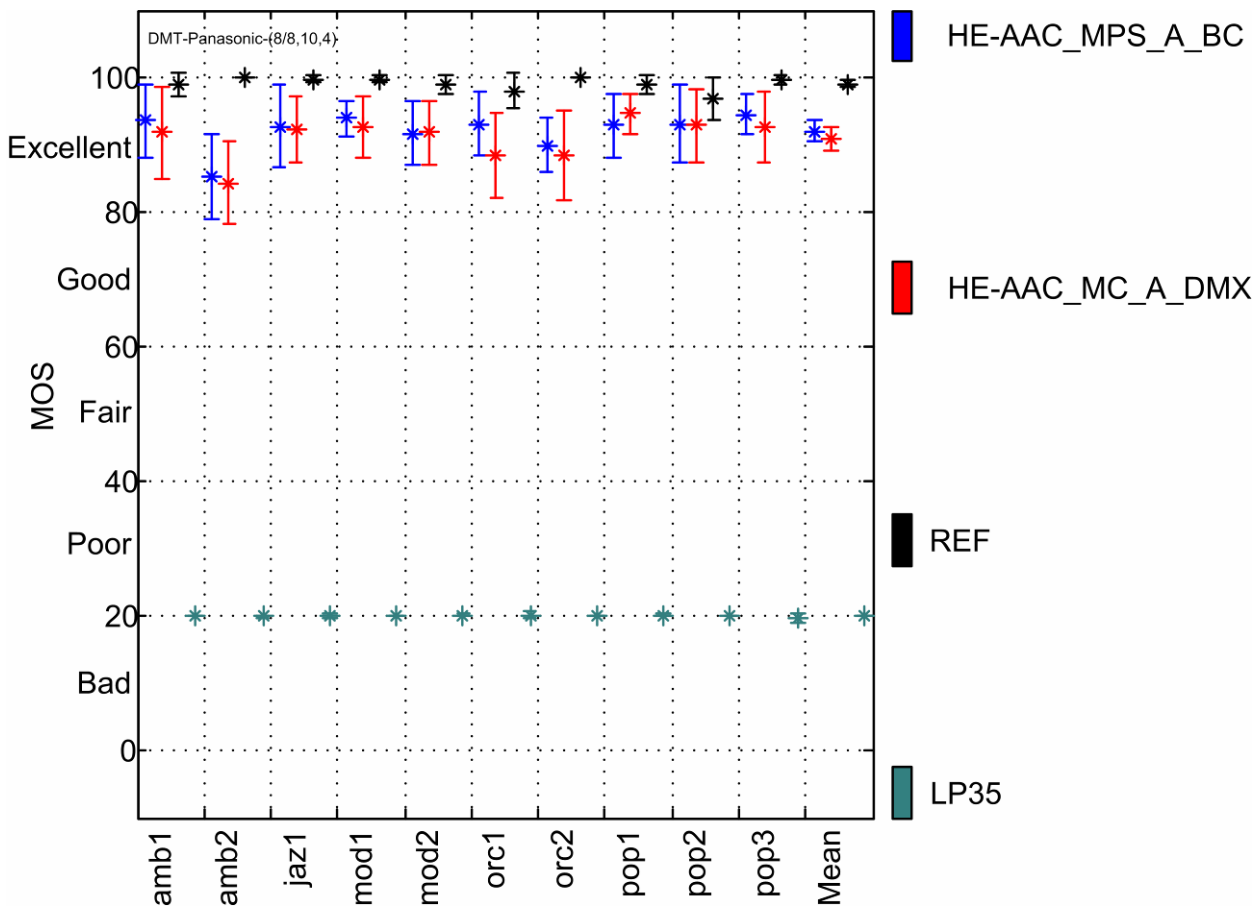

Figure 23 – Results for downmix test, test site Panasonic.

© ISO/IEC 2007 – All rights reserved

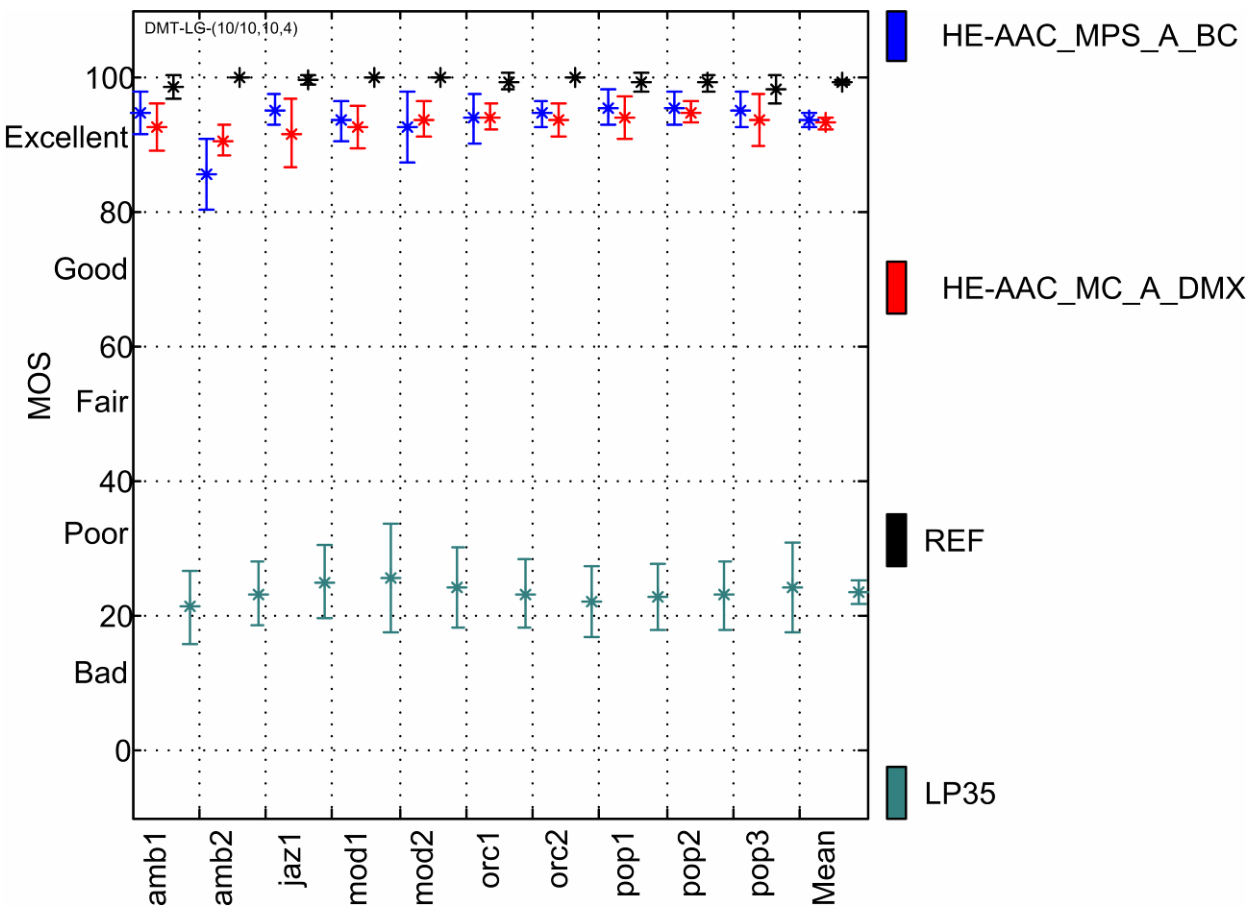

Figure 24 – Results for downmix test, test site LG.

Table 21 - Overall results per test site.

Codec ID	Results per test site, Mean (lower,upper)	
	Panasonic	LG
HE-AAC_MPS_A_BC	92.100 (90.537,93.663)	93.660 (92.489,94.831)
HE-AAC_MC_A_DMX	91.013 (89.257,92.768)	93.180 (92.210,94.150)
REF	99.100 (98.589,99.611)	99.470 (99.117,99.823)
LP35	19.988 (19.880,20.095)	23.500 (21.755,25.245)

A.7.3 Multi-channel test 2

This section provides both the overall results as well as the results on a per item basis per test site for multi-channel test 2. In addition also the mean and confidence intervals are tabulated.

© ISO/IEC 2007 – All rights reserved

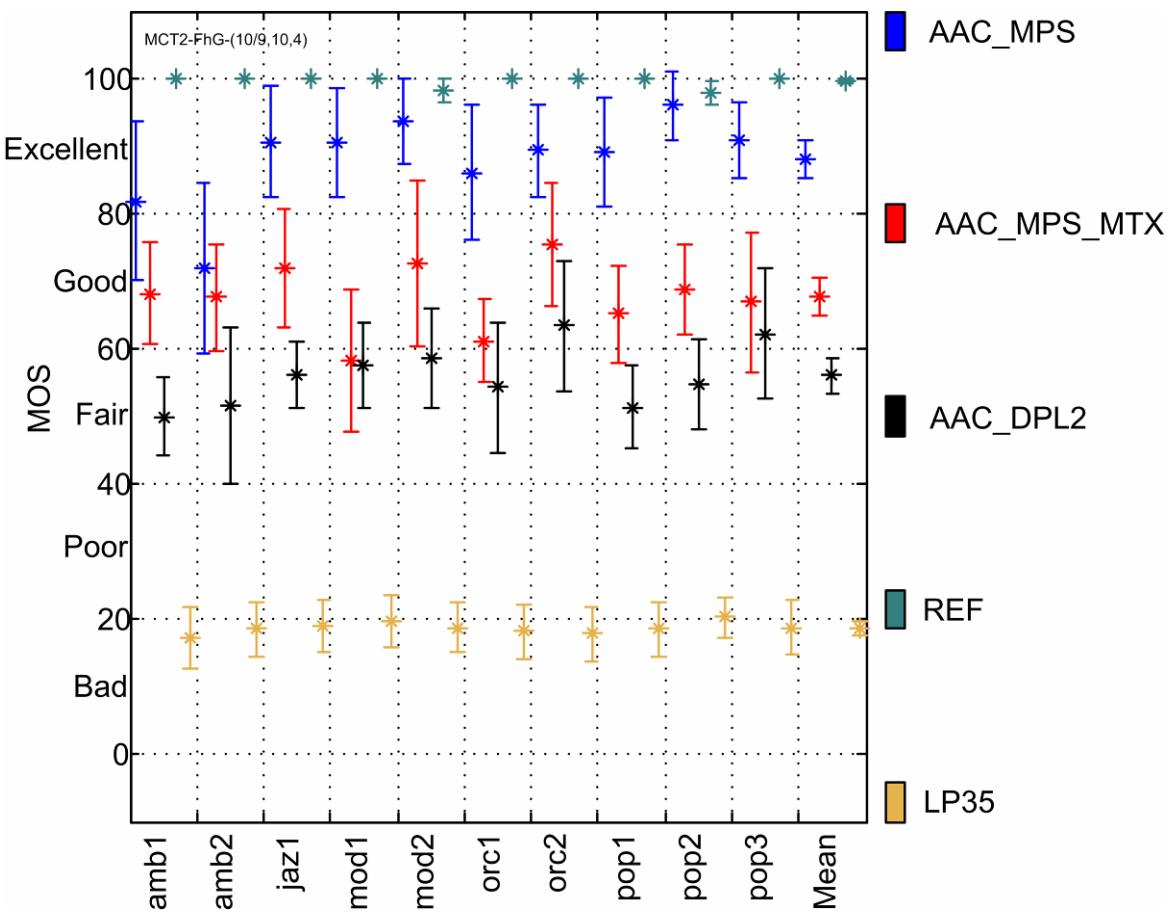

Figure 25 – Results for multi-channel test 2, test site FhG-IIS.

© ISO/IEC 2007 – All rights reserved

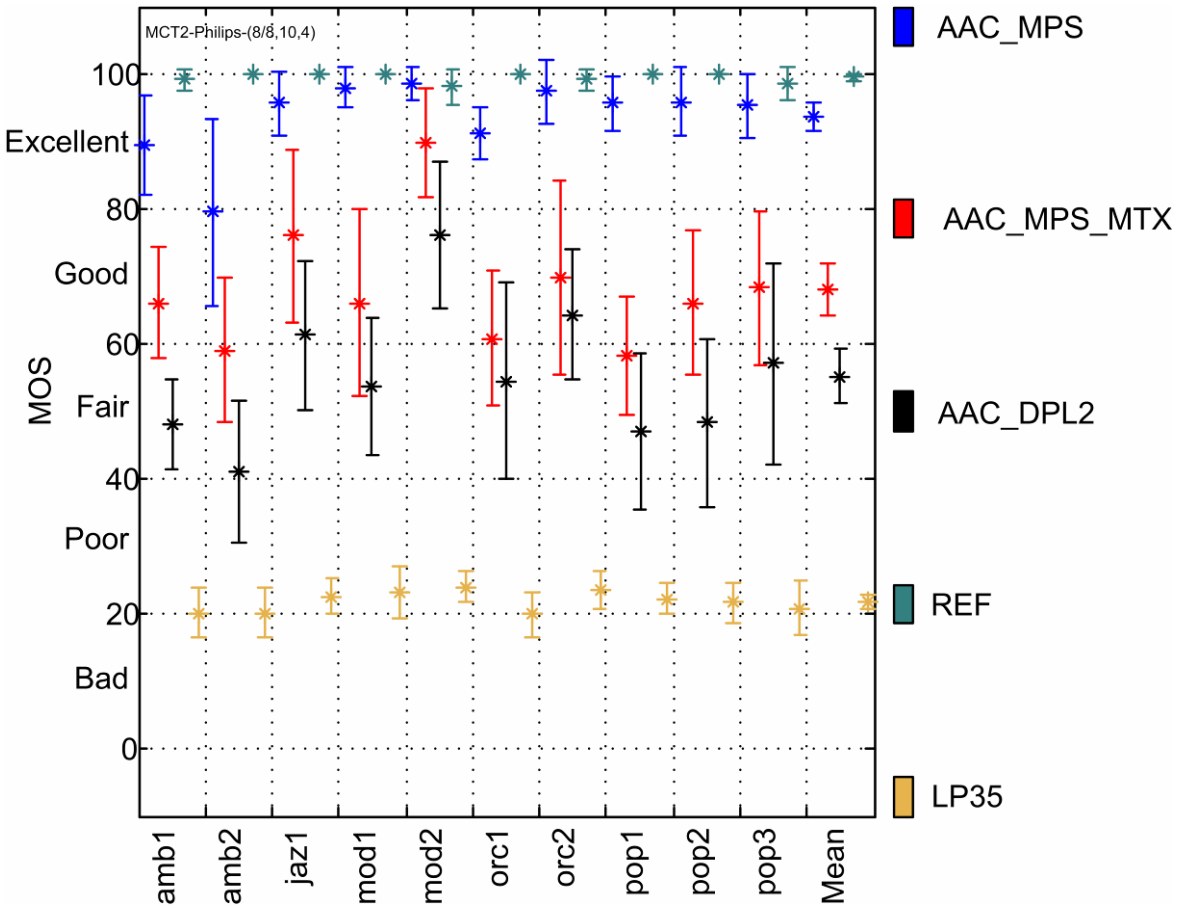

Figure 26 – Results for multi-channel test 2, test site CT.

© ISO/IEC 2007 – All rights reserved

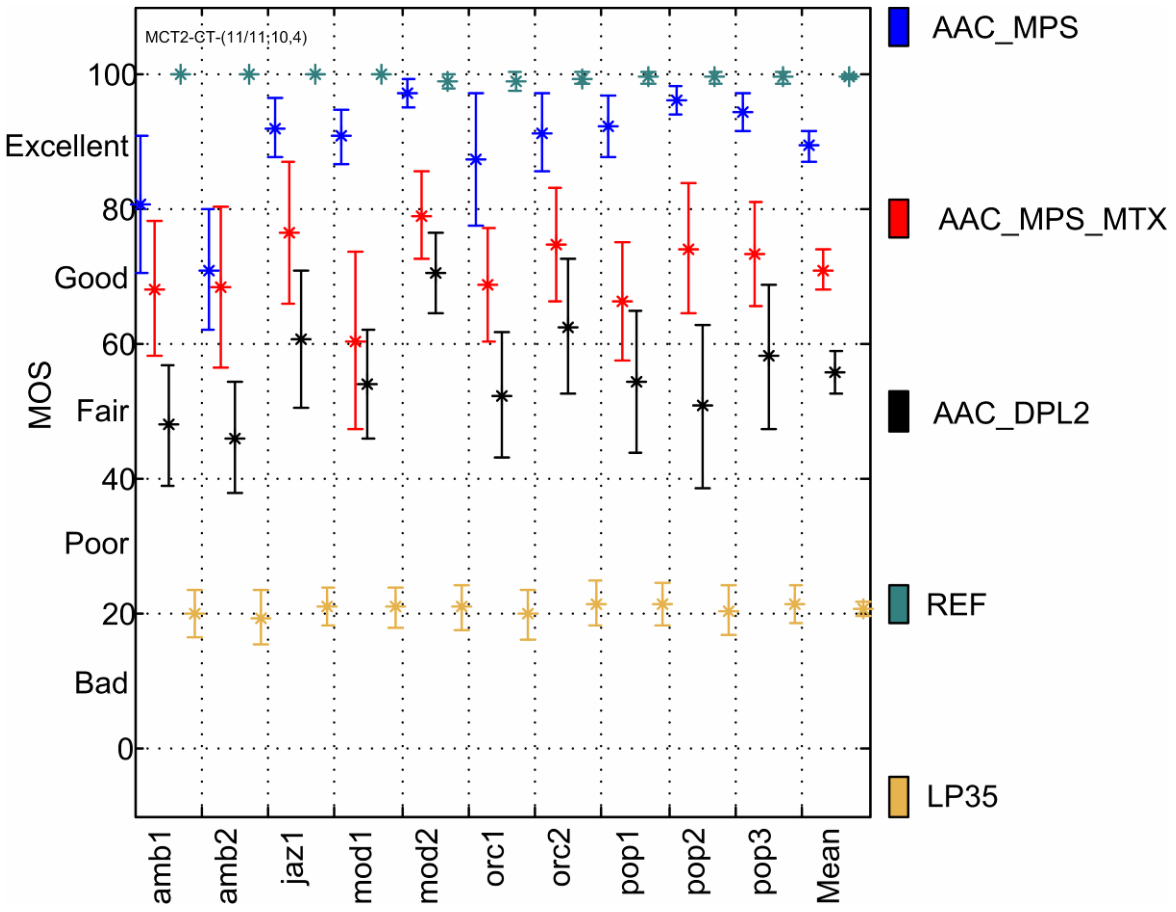

Figure 27 – Results for multi-channel test 2, test site Thomson Hannover.

© ISO/IEC 2007 – All rights reserved

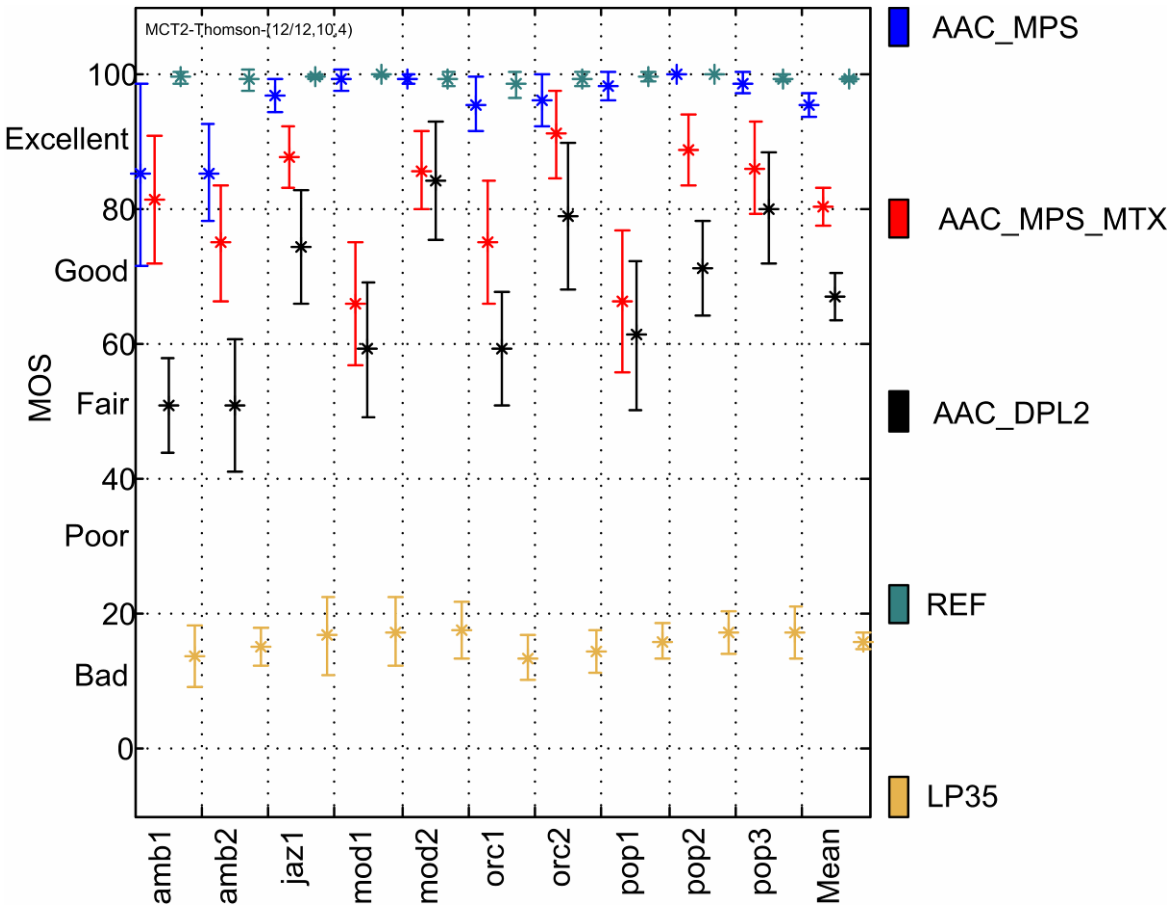

Figure 28 – Results for multi-channel test 2, test site ETRI.

Table 22 - Overall results per test site.

Codec ID	Results per test site, Mean (lower,upper)			
	FhG-IIS	Philips	CT	Thomson Hannover
AAC_MPS	88.000 (85.098,90.902)	93.700 (91.495,95.905)	89.300 (86.946 ,91.654)	95.475 (93.610,97.340)
AAC_MPS_MTX	67.600 (64.755,70.445)	68.025 (64.191,71.859)	71.036 (67.938, 74.135)	80.292 (77.448,83.135)
AAC_DPL2	56.000 (53.445,58.555)	55.163 (51.134,59.191)	55.745 (52.567, 58.924)	67.033 (63.563,70.503)
REF	99.622 (99.339,99.905)	99.513 (99.083,99.942)	99.600 (99.353, 99.847)	99.408 (99.087,99.729)
LP35	18.622 (17.423,19.821)	21.813 (20.795,22.830)	20.700 (19.679, 21.721)	15.850 (14.604,17.096)

A.7.4 Binaural test

This section provides both the overall results as well as the results on a per item basis per test site for the binaural test. In addition also the mean and confidence intervals are tabulated.

© ISO/IEC 2007 – All rights reserved

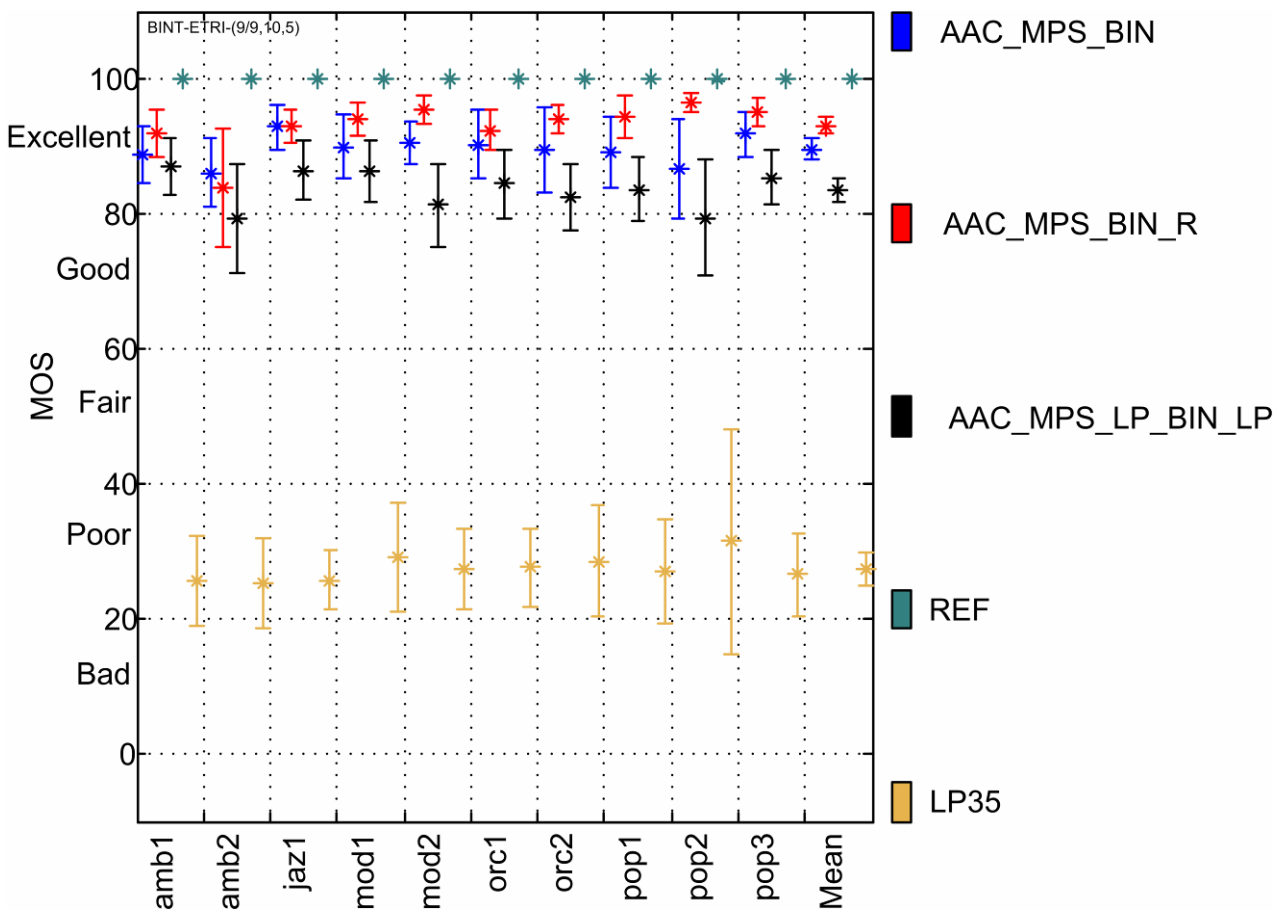

Figure 29 – Results for binaural test, test site ETRI.

© ISO/IEC 2007 – All rights reserved

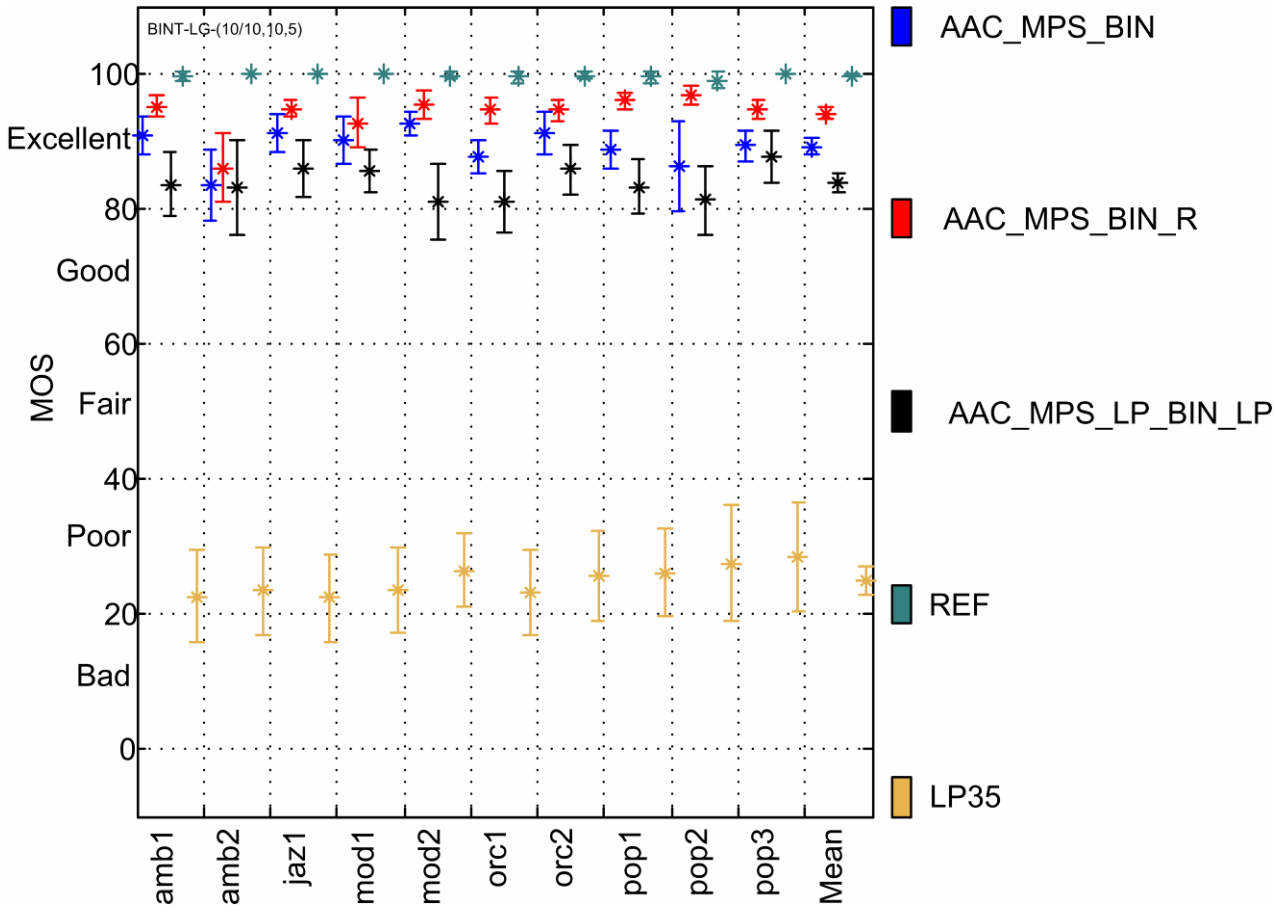

Figure 30 – Results for binaural test, test site LG.

© ISO/IEC 2007 – All rights reserved

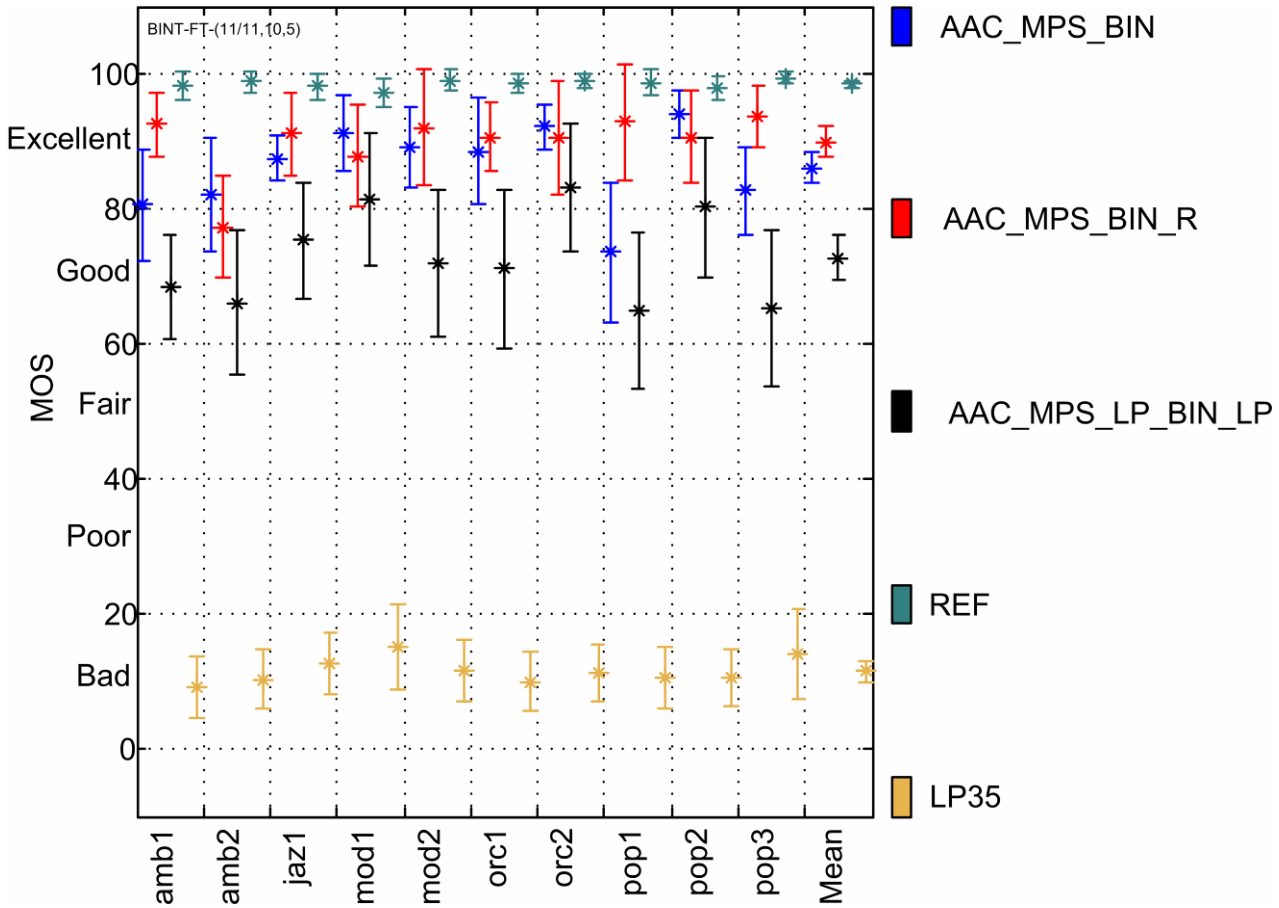

Figure 31 – Results for binaural test, test site FT.

Table 23 – Overall results per test site.

Codec ID	Results per test site, Mean (lower,upper)		
	ETRI	LG	FT
AAC_MPS_BIN	89.533 (87.992,91.074)	89.210 (88.007,90.413)	86.136 (83.817,88.456)
AAC_MPS_BIN_R	93.078 (91.770,94.386)	94.100 (93.177,95.023)	89.936 (87.668,92.204)
AAC_MPS_LP_BIN_LP	83.567 (81.772,85.361)	83.870 (82.385,85.355)	72.791 (69.420,76.162)
REF	99.989 (99.967,100.011)	99.730 (99.526,99.934)	98.491 (97.970,99.012)
LP35	27.433 (24.921,29.946)	24.910 (22.829,26.991)	11.500 (9.976,13.024)

A.7.5 Technology test

This section provides both the overall results as well as the results on a per item basis per test site for the technology test. In addition also the mean and confidence intervals are tabulated.

© ISO/IEC 2007 – All rights reserved

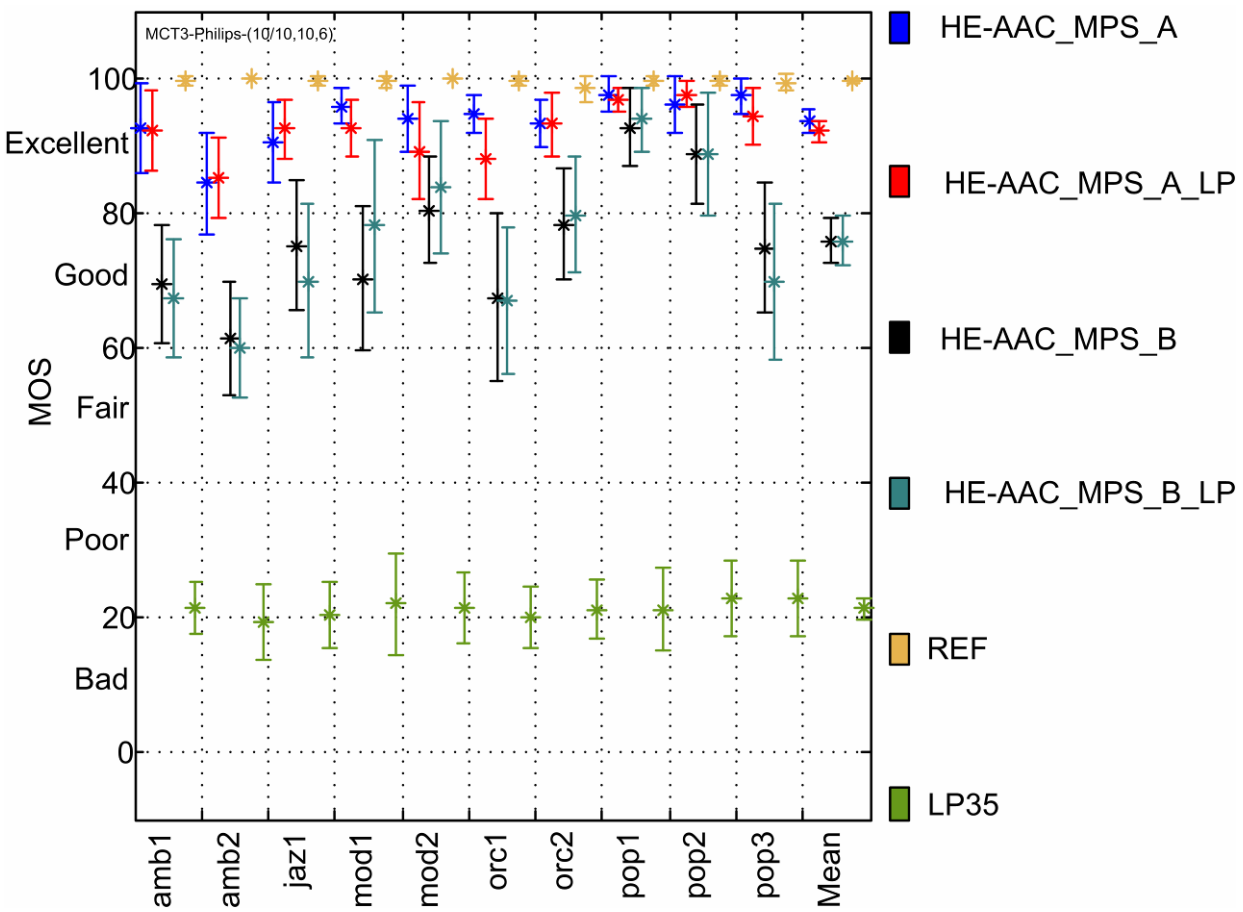

Figure 32 – Results for multi-channel test 3, test site Philips.

© ISO/IEC 2007 – All rights reserved

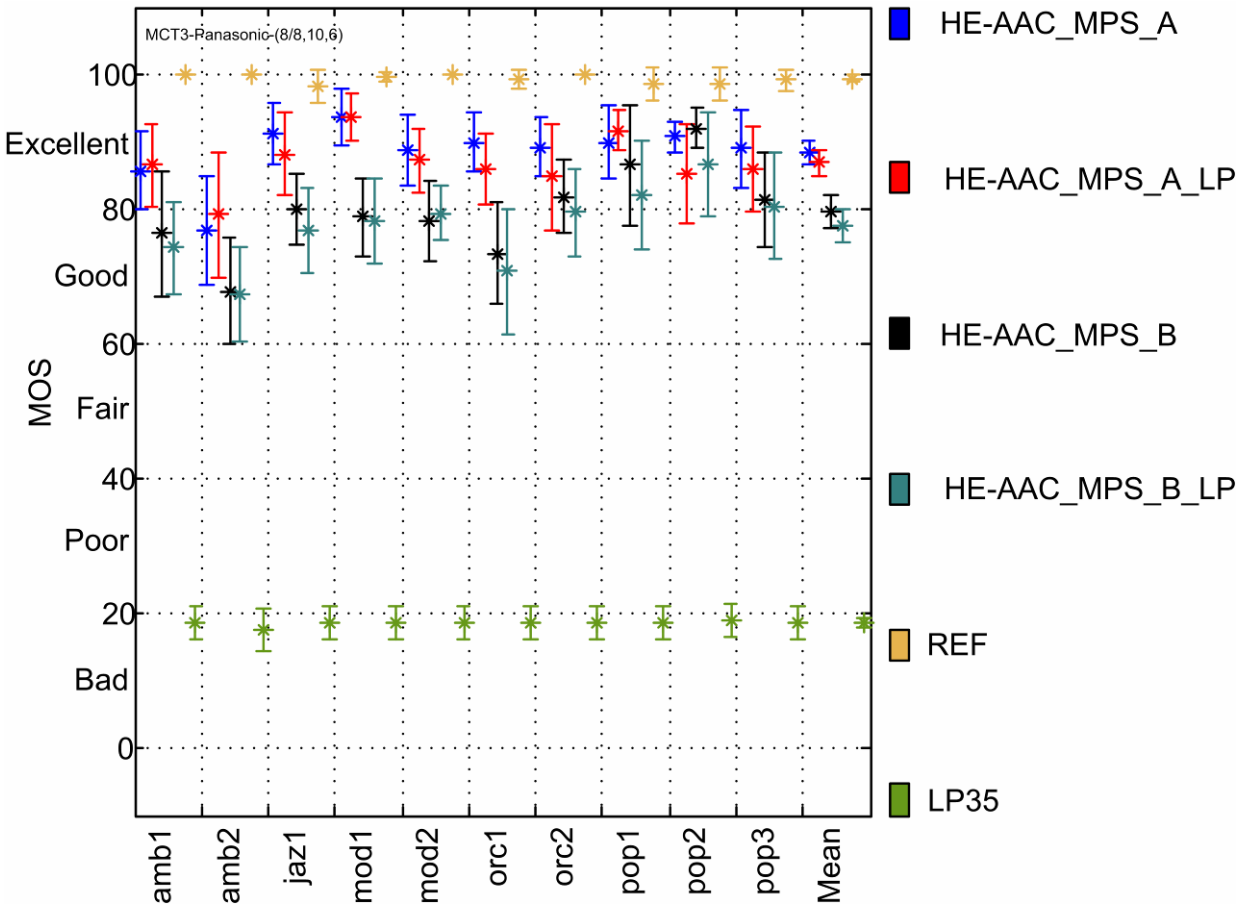

Figure 33 – Results for multi-channel test 3, test site Panasonic.

Table 24 - Overall results per test site.

Codec ID	Results per test site, Mean (lower,upper)	
	Philips	Panasonic
HE-AAC_MPS_A	93.680 (92.086,95.274)	88.487 (86.649,90.326)
HE-AAC_MPS_A_LP	92.200 (90.544,93.856)	86.850 (84.816,88.884)
HE-AAC_MPS_B	75.920 (72.640,79.200)	79.662 (77.190,82.135)
HE-AAC_MPS_B_LP	75.910 (72.338,79.482)	77.563 (75.135,79.990)
REF	99.550 (99.251,99.849)	99.388 (98.911,99.864)
LP35	21.260 (19.609,22.911)	18.550 (17.793,19.307)

© ISO/IEC 2007 – All rights reserved