BBC

R&D White Paper

WHP 025

March 2002

Report on formal subjective viewing tests of MPEG-2 video encoding for High Definition and Standard Definition Television on Plasma Display Panels

T. Davies ***and*** **J.O. Drewery**

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 025

Report on formal subjective viewing tests of MPEG-2 video encoding for High Definition and Standard Definition Television on Plasma Display Panels

Thomas Davies and John Drewery

Abstract

Tests have been conducted on MPEG-2 coded video derived from material scanned at 720p (HD) and displayed on 480-line and HD plasma displays. The objects of the tests were twofold; firstly, to discover whether, for HD display, there was any advantage, over straight coding, in coding the 720p material in 576i form and up-converting back to the HD display's format; secondly, whether, for 480-line display, there was any advanatage in coding before, rather than after, conversion to 576i format and then to the display's format.

The EBU Double Stimulus Continuous Quality (DSCQ) method was used to measure the difference between an uncoded reference signal and signals coded at various bit rates. Results show that, for HD display, it is neceassary to use 19 Mb/s to ensure that pictures delivered by HD are significantly better than when delivered by 576i. However, there was great variation between picture material and whilst some pictures were better delivered via 576i, others were better delivered via HD at all rates measured. For delivery to a 480-line display, 8-10 Mb/s is required to produce results within half a grade of the uncoded picture. Some picture sequences were slightly better delivered via coding at HD at bit rates over 6.4 Mb/s.

This document was originally presented to the EBU Displays Group Meeting at BBC Research and Development, Kingswood Warren, 25-26 March 2002.

Key words: MPEG-2 coding, 720p, HD compression, Plasma Displays

White Papers are distributed freely on request.

Authorisation of the Head of Research is required for publication.

BBC Research & Development
White Paper WHP 025

Report on formal subjective viewing tests of MPEG-2 video encoding for High Definition and Standard Definition Television on Plasma Display Panels

Thomas Davies and John Drewery

1 Introduction

In October 2000, Plasma Display Panel (PDP) tests were conducted in Turin [1] to establish the effect of bit rate on the subjective quality on large PDPs. These displays are being purchased in increasing numbers, and the Turin tests showed that they exposed compression artefacts introduced by MPEG-2 coding which were invisible or less significant on CRT displays. In order to maintain quality within 1 point of the original pictures, it was determined that at least 7-8 Mb/s was required, implying that a review of transmission bit rates may be necessary. Subsequently, it was agreed within the EBU that further tests were required to investigate delivery via High Definition Television (HDTV) standards as well as Standard Definition Television (SDTV) to large PDPs, and BBC R&D undertook to perform these tests at Kingswood Warren. These tests were conducted in January and February 2002 and this document describes the results.

There were two main aims of the tests. The first was to repeat the Turin test for an HD standard, and thus find the MPEG-2 bit rate at which HDTV can be delivered with good quality. The second aim was to compare SDTV delivery with HDTV delivery on large PDPs, and determine at what point HDTV delivery becomes preferred. The tests that were conducted therefore compared both HD and SD material, coded at various bit rates, on PDPs with both SDTV and HDTV interfaces against both HD and SD uncompressed references. As a result we were able to address both these questions, and in addition independently check the results of the Turin tests.

What lay behind these aims was a realisation that developments in display technology, encompassing domestic projection systems as well as PDPs, have created a problem and an opportunity for broadcasters. The problem, exposed by the Turin tests, is that conventional digital Standard Definition broadcasts at normal bit rates suffer considerably when displayed on large PDPs. The opportunity presented by these PDPs is the creation of a market for enhanced quality television delivery, already fuelled by SDTV DVDs. This market will consist of consumers with displays of greatly varying size and resolution. It is therefore not necessarily the case that it will be best served by the delivery of very high bit rate services at very high resolution. Instead it is possible that viewers could be satisfied with services that offer a significant enhancement over current SDTV broadcasts, but which might fall well short of the highest quality HDTV. A key question therefore, is whether in these circumstances it is better to increase SDTV bit rates to nearer DVD levels, or to jump to a new delivery standard.

An additional factor that influenced the design of the tests was the existence of the 720p standards[1]. These standards have a number of advantages in the context of these tests. Firstly, as progressive standards they are matched to progressive-scanned displays such as PDPs, and so standards-conversion artefacts are reduced. Secondly, it is generally thought that ‘progressive video is easier

[1] 720p refers to a number of standards, each progressive-scanned with 720 lines x 1280 pixels per frame. The frame rates that have been standardised are 24, 50, 60 and 59.94. Throughout this document we will refer to 720/50p as 720p for convenience.

to compress than interlaced'. What this actually means and for what spatial and temporal resolutions and picture material this might be true, is a complex issue but there is certainly some truth in it. Thirdly, 720p represents a more modest increase in spatial resolution over 576i SD than other standards and so is probably more suited to the domestic market. Finally, using 720p at 50 Hz allows for simple down-conversion to 576i, and so both SD and HD material could be derived from a single master and thus directly compared. This would not have been possible if an interlaced HD standard (such as 1080i) was used.

BBC R&D therefore opted to perform tests comparing 720p with 576i delivery in a number of configurations. The use of 720p allowed for a formal subjective assessment methodology to be applied, and the tests were compliant with ITU Recommendation 500 [2].

2 Description of the tests

2.1 The HD Test: 768-line PDP

The configuration of the 768-line PDP test is shown in Figure 1 below:

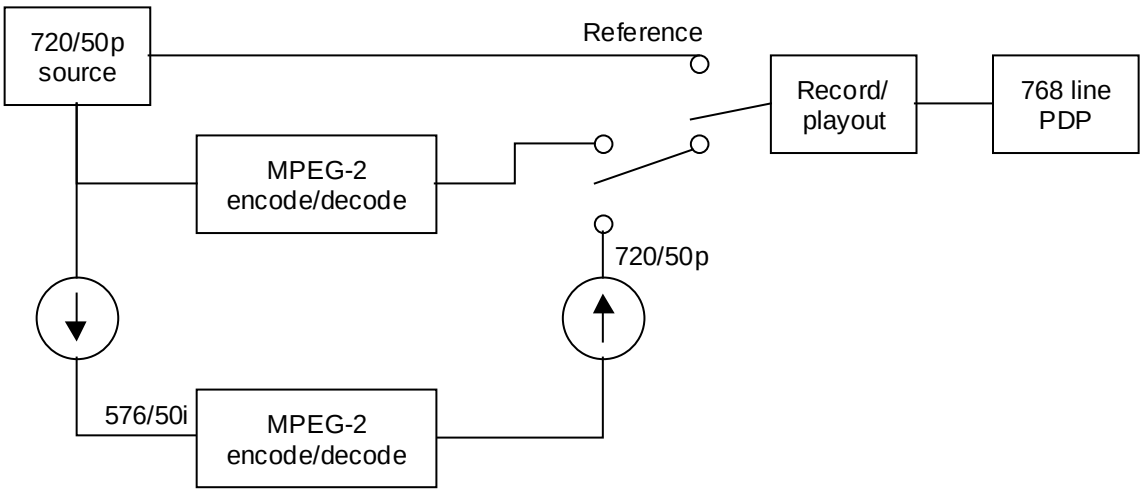


Figure 1: HD Test (768-line PDP) configuration

In this test two sets of sequences are derived from each 720p source. The first (Series 1) is coded at various bit rates at HD before delivery to the panel via HD SDI interfaces. The second (Series 2) is down-converted to SD before coding and is then up-converted to HD and delivered to the panel. The up-conversion step is chosen to use essentially the same filter as is used internally within the panel for conversion from 576i, so that Series 2 represents delivery to the panel via SDI (for more details see section 3.4 below), the delivery of dual standards to the panel being impractical in a subjective test environment. Each test presentation consisted of a comparison of the uncompressed reference – the 720p source in this case – with a member of either Series 1 or Series 2, or (without informing the test subjects) the reference itself.

The HD test provides information on how viewers rate the degradations resulting solely from compression at HD and those resulting from standards conversion and compression at SD. An intriguing part of the test was therefore the effect of interlace-to-progressive conversion artefacts – which were significant on some sequences, notably *tennis* – and the potential benefits of delivering a progressive-scan format to a progressive-scan display technology.

2.2 The SD Test: 480-line PDP

The configuration of the 480-line PDP test is shown in Figure 2 below:

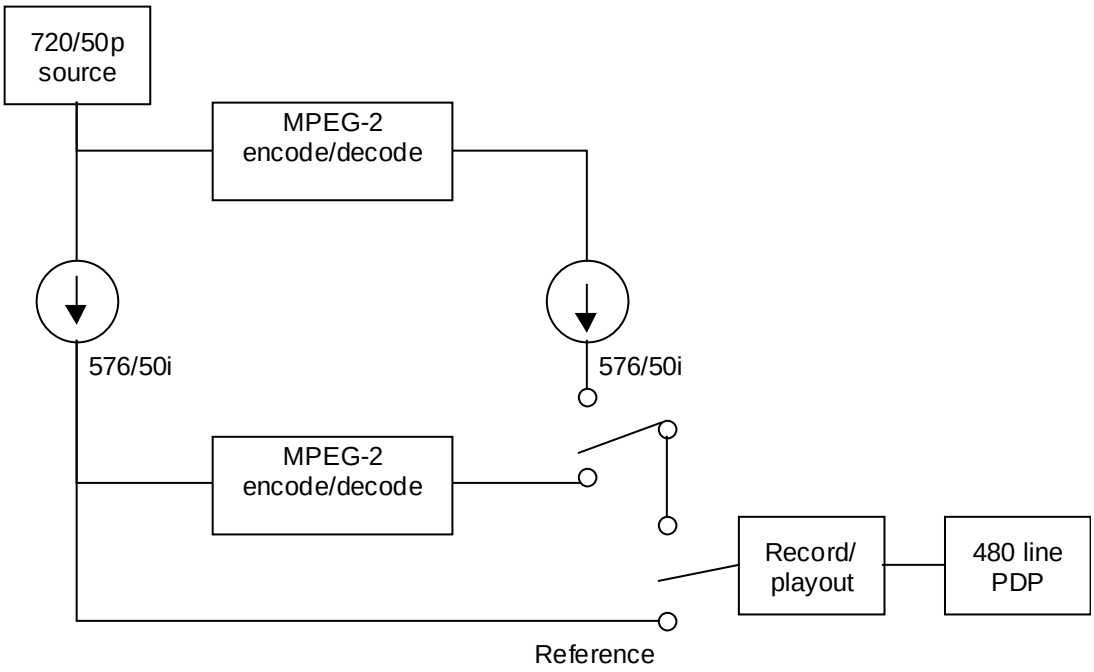


Figure 2: SD Test (480-line PDP) configuration

In this test two sets of sequences are, as for the HD Test, derived from each 720p HD source. The first (Series 1) is down-converted to SD and then MPEG coded at various bit rates before delivery to the panel via an SDI interface. The second (Series 2) is coded at HD first before being down-converted and delivered to the panel, again via SDI. Each presentation to the test subjects again consisted of a comparison between the uncompressed reference and either a member of Series 1, or Series 2, or the reference.

In this test, it was not appropriate (and indeed it was not possible) to use the 720p source as the reference. Instead, as indicated in Figure 2, the down-converted uncompressed HD source was used as the reference.

Whilst it might appear perverse to test HD delivery by means of a display possessing only standard definition interfaces, this test fulfils several functions. Firstly, and most importantly, it tests whether HD delivery disadvantages viewers who cannot experience the greater resolution due to the limitations of their displays, or indeed whether it can provide any benefits for these viewers. Thus it informs the discussion as to whether an HD service should be launched in parallel with SD services, or whether it is more beneficial to provide a single service and incorporate standards conversion into set-top boxes for those whose displays are not HD-capable.

Secondly, the test provides a measure of the efficiency and scalability of delivery via MPEG-2. Although coding 720/50p requires the coding of 4-5 times as many macro-blocks as for 576/50i, the coding of a progressive standard, and other compression factors, may mean that this disadvantage can be mitigated or even overcome. By removing any advantage the HD delivery route might have in terms of additional resolution, the test exposes the efficiency of the compression mechanism.

Finally, as noted in the introduction, the quality and capabilities of viewers' displays varies enormously, and the disparity between display equipment will increase further. The inclusion of an SD display places another point on this continuum.

3 Test methodology

3.1 ITU Rec 500 compliance

General viewing conditions. The following characteristics were measured and observed to be in accordance with the standard:

- Ratio of luminance of inactive screen to peak luminance: PDPs satisfied the limit of < 0.02;
- Ratio of screen luminance at black level and peak white: PDPs satisfied the limit of 0.01;
- Display brightness and contrast: set up by PLUGE;
- Ratio of display background luminance to peak luminance of picture: set to be 0.15 by lighting;
- Chromaticity of background: set to be D65 by curtain colour and lighting;
- Other room illumination: low.

Maximum viewing angle was set by the viewing distance and the number of observers. In the SD test the viewing distance was set at four picture heights ('4H'), whilst in the HD test it was set at three picture heights ('3H'). The choice of viewing distances was influenced by a desire to accord with common practice, in which tests of standard definition pictures are usually carried out at 4H and 6H, whilst 'high definition' is usually understood to be viewed at half the distance for standard definition. However, it was felt that the larger size of the displays reduced the need to test standard definition at 6H and that cell disturbance effects made it uncomfortable to view high definition at 2H.

The use of a smaller number of picture heights for viewing larger displays keeps the ratio of the viewing distance to observer separation comparable with tests on smaller displays. Thus the number of observers is comparable also. If the viewing angle is confined to the prescribed limit of 30 degrees then only 4 observers could be tolerated in each session. Taking into account the actual values for display height and viewer separation, the actual maximum viewing angles were 18 degrees for the SD test and 20 degrees for the HD test.

There is however a further factor that influences viewing angle in the present situation. This is that, as the (16:9) display width is comparable with the distance between observers, the viewing angle varies significantly from one side of the picture to the other and the above values apply only at the centre of the picture. In the SD test the variation, for the edge viewers, is actually from 6 to 29 degrees whilst in the HD test it is from 4 to 33 degrees.

Source signals. These comply with the requirements for type of storage and playout.

Selection of test materials. This is dealt with in section 3.3 below.

Range of conditions and anchoring. Distributed throughout session and not identified.

Observers. 15 observers were used for the SD test and 20 for the HD test. These had a mix of experience in assessing picture material, ranging from novice to experienced. Some had digital receivers in the home and others not. 7 of the observers took part in both tests.

Instructions for the assessment. See Appendix A.

The test session. Rules obeyed on length, dummy presentations, and order randomisation.

Presentation of the results. Means and 95% confidence limits shown. See below.

The double-stimulus continuous quality-scale method. Used with unimpaired reference, either first or second, including reference with itself.

Arrangement. See Figs. 1 and 2 above.

Presentation of the test material. Variant II used, with modifications for captions:

Content	Caption: Test n	A	Grey	B	Grey	Caption: Test n*	A	Grey	B	Grey
Secs.	2	10	2	10	1	2	10	2	10	11

Grading scale. 100 mm continuous scale with qualities: Excellent/Good/Fair/Poor/Bad as per ITU Recommendation 500 [2].

Analysis of the results. Means and 95% confidences of difference from reference. No observers were eliminated, and no score inversion was performed.

3.2 Plasma Display Panels and interfaces

Details of the Plasma Display Panels used are as follows:

	SD Panel	*HD Panel*
Make	Fujitsu	Delphi
Diagonal size, inch	42	50
Aspect ratio	16:9	16:9
Width (*pixels*) x *height* (*lines*)	852x480	1280x768
Glass	4205	503
Grey scale	8-bit native, augmented to 16-bit	12-bit native, augmented to 16-bit
Contrast ratio	70:1	90:1
Refresh rate, Hz	50/60	50/60
Display rate, Hz	100/120	50/60
Input	SDI	HDSDI and SDI

3.3 Origin and selection of material

Five video sequences were selected as source material for the tests. The twin aims in selecting material was to provide a range of compression difficulty (although biased towards challenging material) and a wide range of subject matter. Five sequences were eventually selected from a variety of sources, together with a sixth for use as training material. These were:

Waterfall. This sequence showed Yosemite Falls, and was originally shot on film. The sequence had been scanned to 1080/24p, and for the tests was down-converted to 720/24p and played out at 50Hz instead of 24: since the original sequence was overcranked this produced a believable sequence. The sequence was typical of high quality natural history programming that would benefit greatly from HD presentation. It combined slow horizontal motion with high levels of texture, a combination particularly challenging for compression in an interlaced format.

Panslow. This sequence was a studio shot selected from a number of sequences created by an EBU team in Turin in October 2000 using an LDK6000 camera, running at a 720/50p standard and transferred to disk without compression. The sequence was transferred to Kingswood via HD D5, and so underwent slight compression before processing. It consists of a slow crab across a studio

scene comprising several items, thought to be critical for testing resolution impairment and creating the sort of noise added by MPEG coding.

Swim-reflection. This sequence was originated by a Panasonic 27V (variable frame rate) camera, running at a 720/60p standard and recorded on the camera's tape using the DVCproHD standard which has a compression of 6 or 7:1, including 1280 – 960 horizontal conversion. This signal was then transferred to a RAM store at Kingswood using a Panasonic special box, before transferring to computer for processing. The content comprises a bird swimming on rippling water amidst highly coloured reflections of the surrounding environment. Despite limited spatial resolution, the subtle water motion and the presence of specular reflections suggested the sequence would be very challenging.

Tennis. This sequence was originated in the same way as *swim-reflection*. The content comprises a medium shot of a tennis player returning shots with a scoreboard and crowd behind. The player provides plenty of motion, augmented by slow diagonal camera movement. The sequence is not particularly testing in resolution terms, but contains a number of near-horizontal features such as railings, thereby exercising the conversion to and from interlace.

Watercomp. This sequence was originated in the same way as *swim-reflection*. The content is a cross-fade between two sequences, one showing water bubbling vigorously into a pool and the other a close-up of the same pool. Despite there being limited spatial detail, the sequence combines considerable turbulence, specular reflections and a cross-fade, and was felt to be very challenging.

Spincalendar. This was the training sequence. It was originated in the same was as *panslow*, and consists of a slowly rotating board on which a calendar, and other resolution-testing features were mounted.

3.4 Standards conversion

As Figures 1 and 2 show, standards conversion is performed in the tests to convert between 576i and 720p. This was in addition to standards conversion within the display. All of these conversions were done in software.

Down-conversion from 720p to 576i was done in two stages: horizontal and vertical scaling to 576/50p followed by pre-filtering and vertical sub-sampling to create an interlaced picture. The pre-filter selected was the HHI pre-filter, which is less severe than the line averaging that would be performed in a SD camera CCD. This provided a good compromise between reduced line twitter and exploitation of the PDPs' size and resolution.

Up-conversion from 576i to 720p was performed only in the HD test, and was used only for the convenience of the tests since the 50" PDP has both SD and HD SDI interfaces: it avoided switching standards within the test, which would have caused the display to reset. The up-conversion filter therefore used the same 3-field linear filter that is used within the 50" PDP. The PDP up-conversion was not perfectly replicated, however, since the 50" PDP had 768 lines instead of 720 lines and so there was additional vertical scaling in the display.[2] This was of no consequence for the test as the effect of the additional scaling was negligible and it applied equally to both SD- and HD-coded material. In fact, the 720 to 768 scaling was only visible as patterning on the HD training sequence, spincalendar.

[2] It had been hoped to use a 60" 1280 x 720 line PDP for the tests, but it could not be prepared in time.

3.5 MPEG-2 encoding and bitrates

All pictures were encoded and decoded using the SSG reference codec, which is available online [3]. The SSG coder attempts all macro-block modes for each macro-block in order to determine the most efficient coding strategy. Motion vectors are determined by area search with search parameters specified in the parameter file, and proportional to the distance between reference and predicted frames. No attempt to smooth vector fields or trade off vector bit rate against prediction error is made. This means that, whilst the SSG coder is close to an optimum MPEG-2 implementation for high bit rates or unchallenging material, it is sub-optimal for low bit rates or challenging material.

The coding parameters were selected to be optimal for the bit-rate and standard to be encoded, and to be comparable between 720p and 576i encoding where possible.

576i encoding used 4:2:0 MP@ML coding with a standard 12,3 GOP structure. Motion estimation was performed on an area of 18 x 13 pixels per frame period, which covered most of the motion in the selected sequences. The coder also used the alternate scan for DCT coefficients, which in general has slightly better performance for interlaced material.

720p material was encoded using 4:2:0 MP@ML or MP@HL with a 24,6 GOP structure so that the GOP occupied precisely the same time period as that for SD material, ensuring that temporal artefacts would be comparable. Motion estimation was performed on an area of 16×8 pixels per frame period. Taking into account the relative dimensions of the 720p and 576i frames, and the fact that there are 50, not 25, frames per second in the HD case this means that this motion vector search area corresponds almost exactly to the area searched at SD.[3] The coder used the standard scan for DCT coefficients.

All the SD material was coded at 4.8, 6.4, 8, 10, and 14 Mb/s before down-selection for each test. Likewise, with the exception of watercomp, all the HD material was coded at 6.4, 8, 10, 14 and 19Mb/s. It was not possible to code watercomp at 6.4Mb/s at 720p, and so this configuration was eliminated from both tests.

For both HD and SD tests, all the material was reviewed on the relevant PDP at the test distances (4H for the SD test and 3H for the HD test) before down-selection. In fact, the same bit rates were chosen in both tests: material coded at 576i was coded at 4.8, 6.4, 8 and 10 Mb/s, with watercomp also coded at 14Mb/s; material coded at 720p was coded at 6.4, 8, 10 and 19Mb/s, with the exception of watercomp which was coded at 14Mb/s but not at 6.4Mb/s. The anomaly for the watercomp bitrates arose from its exceptionally challenging nature: in particular, it proved impossible to code at HD at a bitrate as low as 6.4Mb/s.[4]

There was considerable discussion in selecting these bit rates, with a number of factors being taken into account. 4.8Mb/s was included because it represents DTT transmission bit rates: 4.8Mb/s CBR is the rate at which BBC1 is broadcast in the BBC DTT multiplex. 10 Mb/s was reluctantly accepted as the maximum SD bit rate for all sequences except watercomp because this is the maximum rate that current DVDs are recorded at. 19Mb/s was included for HD because current HDTV broadcasts are at this rate. Other rates were included to provide a good spread of quality.

---

[3] An area of 16 x 8 every 1/50th sec corresponds to 32 x 16 every 1/25th sec. 32.(702/1280)=17.55 and 16.(576/720)=12.8.

[4] As noted, the SSG coder is suboptimal at low bit rates or for very challenging material. It was possible to code watercomp at HD at 6.4Mb/s using the more sophisticated COUGAR coder developed by BBC R&D and Snell and Wilcox.

4 Analysis and results

4.1 General

This section presents the results of the tests in graphical form, averaged over the participants in the tests. A 1-sided t-test with a 95% confidence level was employed to determine those participants who could not, with 95% confidence, distinguish between coded and reference pictures: all participants passed in both tests.

It will be noted that for a number of test conditions, the confidence intervals are much larger than those of other conditions in the same series. This appears to be almost certainly due to the presence of transcription errors (transposing A and B) made by the subjects, which were not spread evenly over all test conditions. Although it might be possible to apply suitable tests to identify and eliminate such errors, it was decided, after much discussion not to do so as this might confer spurious confidence on the results for the affected condition. The impact of these errors is to reduce confidence in the mean values recorded, and this is reflected in the confidence intervals.

4.2 768 line PDP results

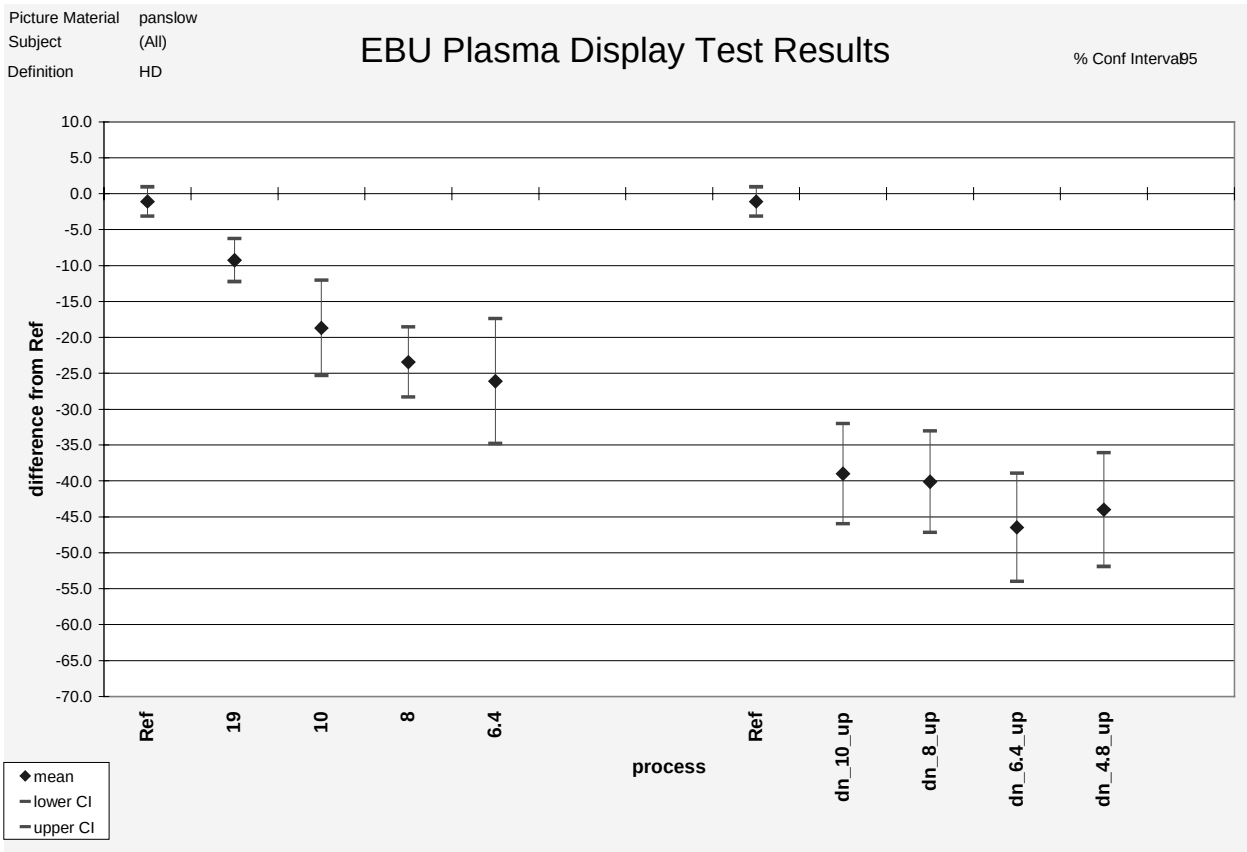
The results of the HD test, involving the 768 line PDP, are shown in Figs. 3 – 10. These show means and 95% confidence limits of the difference between the score for the reference and the score for each processing condition, measured on the 100-point scale. It must be remembered that, on that scale, 20 points corresponds to a single quality grade on a five point scale. The conditions for straight coding, (HD delivery), are grouped before the conditions for down-conversion, coding and up-conversion (SD delivery). Note that the score for reference versus reference is also shown as part of both groupings. As the results for the pictures vary widely, results are shown for each picture separately as well as results for averages.

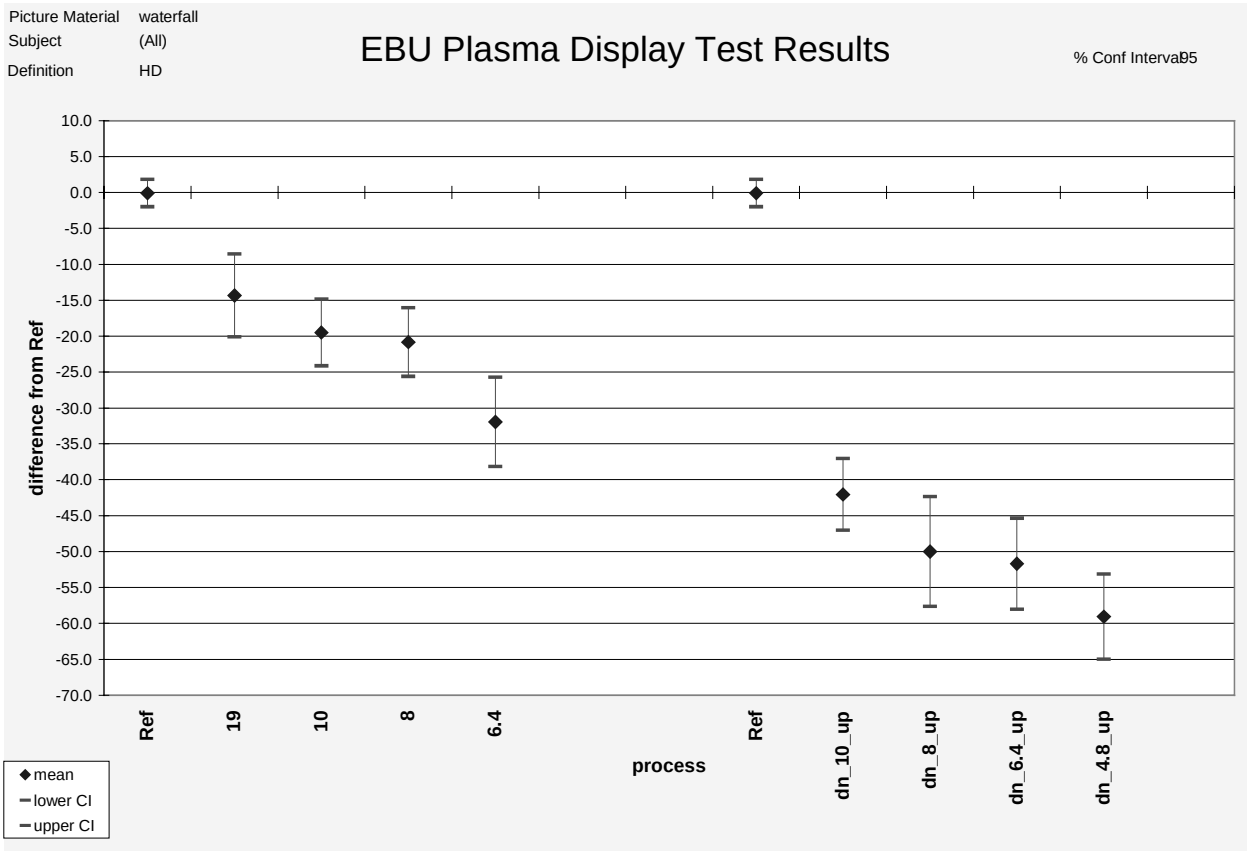
The results for the pictures *panslow*, *waterfall* and *tennis* are markedly different from those for *swim-reflection* and *watercomp*. So the first set have been termed ‘Group 0’ and the second set have been termed ‘Group 1’. In the first case, the mean score for HD delivery at any of the bit rates tested is higher than the mean score for SD delivery at any of the bit rates tested - in almost all cases, the difference exceeds the 95% confidence limits. The trend is shown more clearly by the average over these three pictures. In fact, the asymptotic quality of SD delivery, at infinite bit rate, is probably below that for HD delivery at a rate as low as 6.4Mb/s.

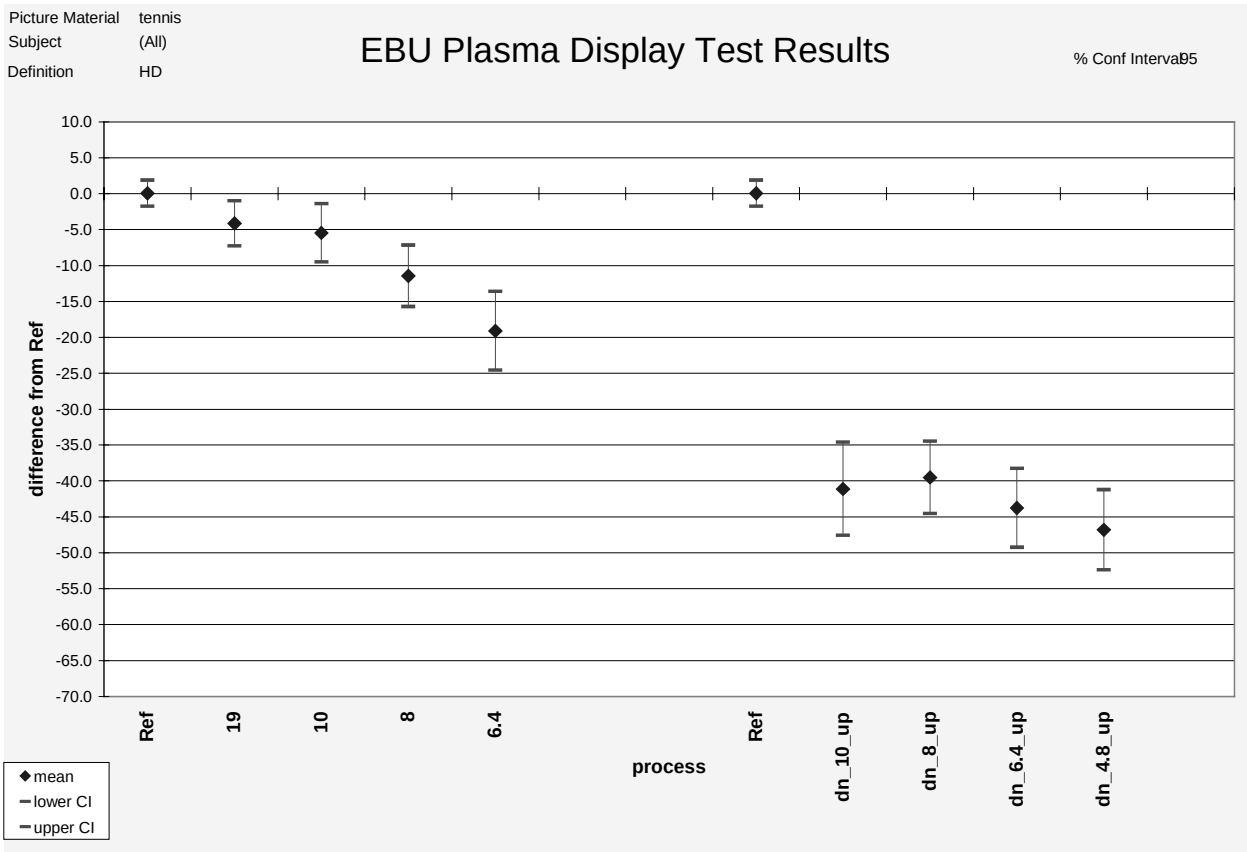
On the other hand, for the pictures *swim-reflection* and *watercomp*, in Group 1, the ‘curve’ for HD delivery is much steeper and clearly crosses the asymptotic value for SD delivery. At lower bit rates, SD delivery is preferred. This is, again shown more clearly by the average for these two pictures.

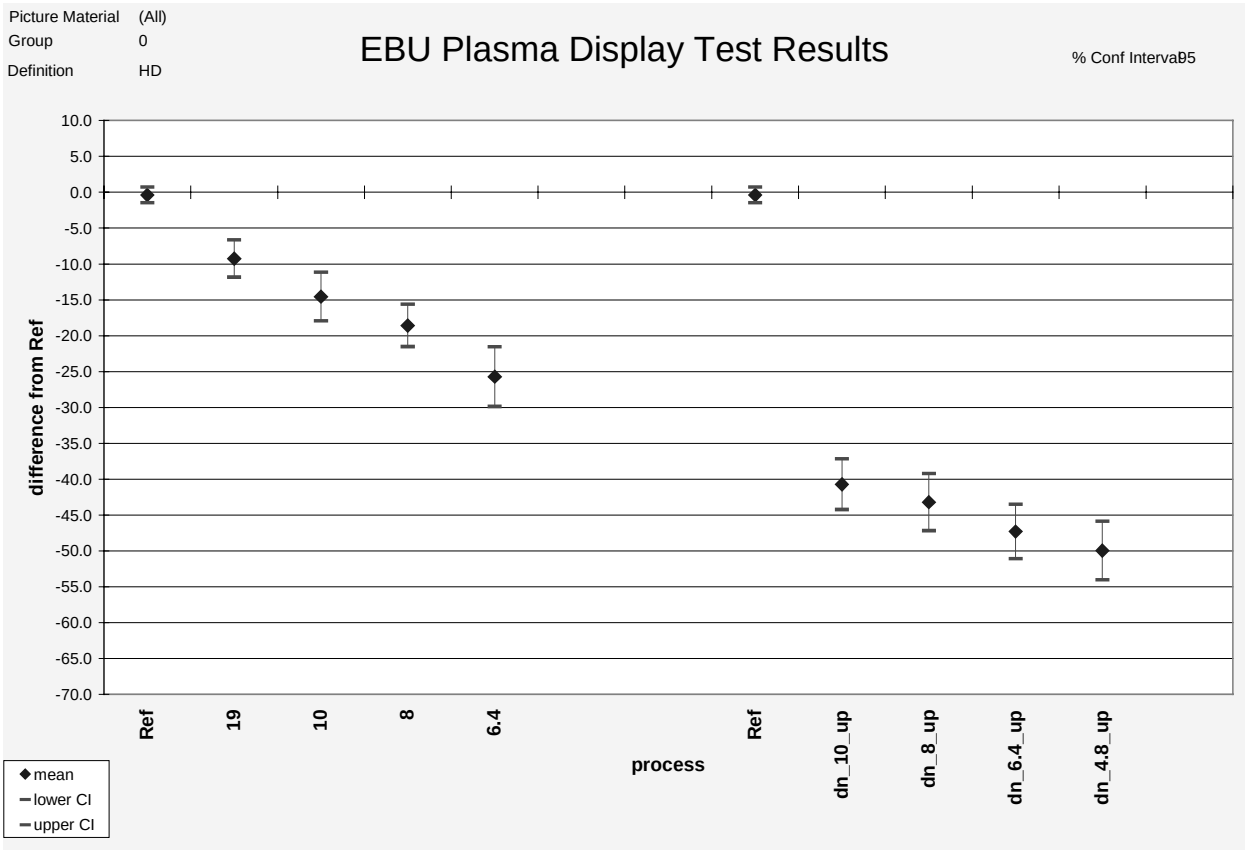
The average for both groups has more of the characteristics of the first group than the second.

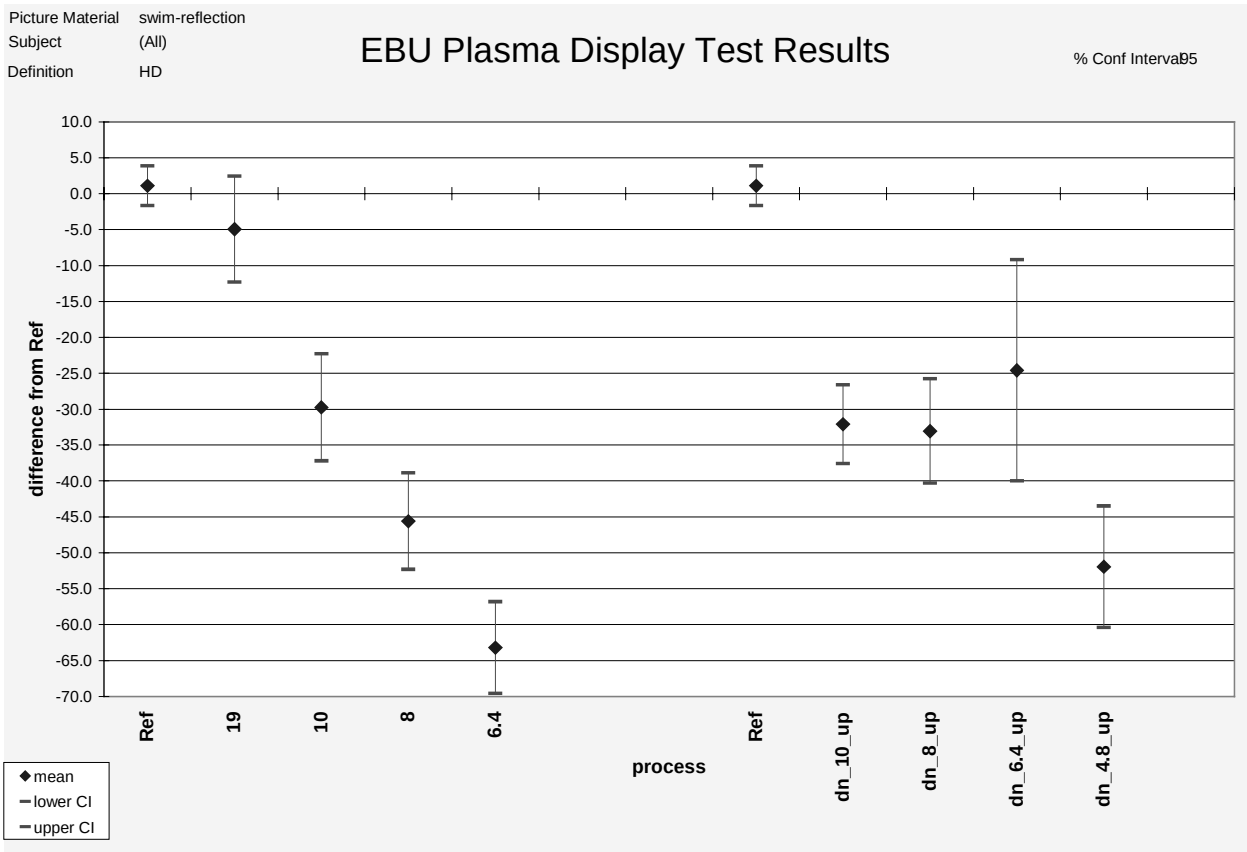
Figs. 3 – 10, following: Results of the HD test for individual pictures and for group averages

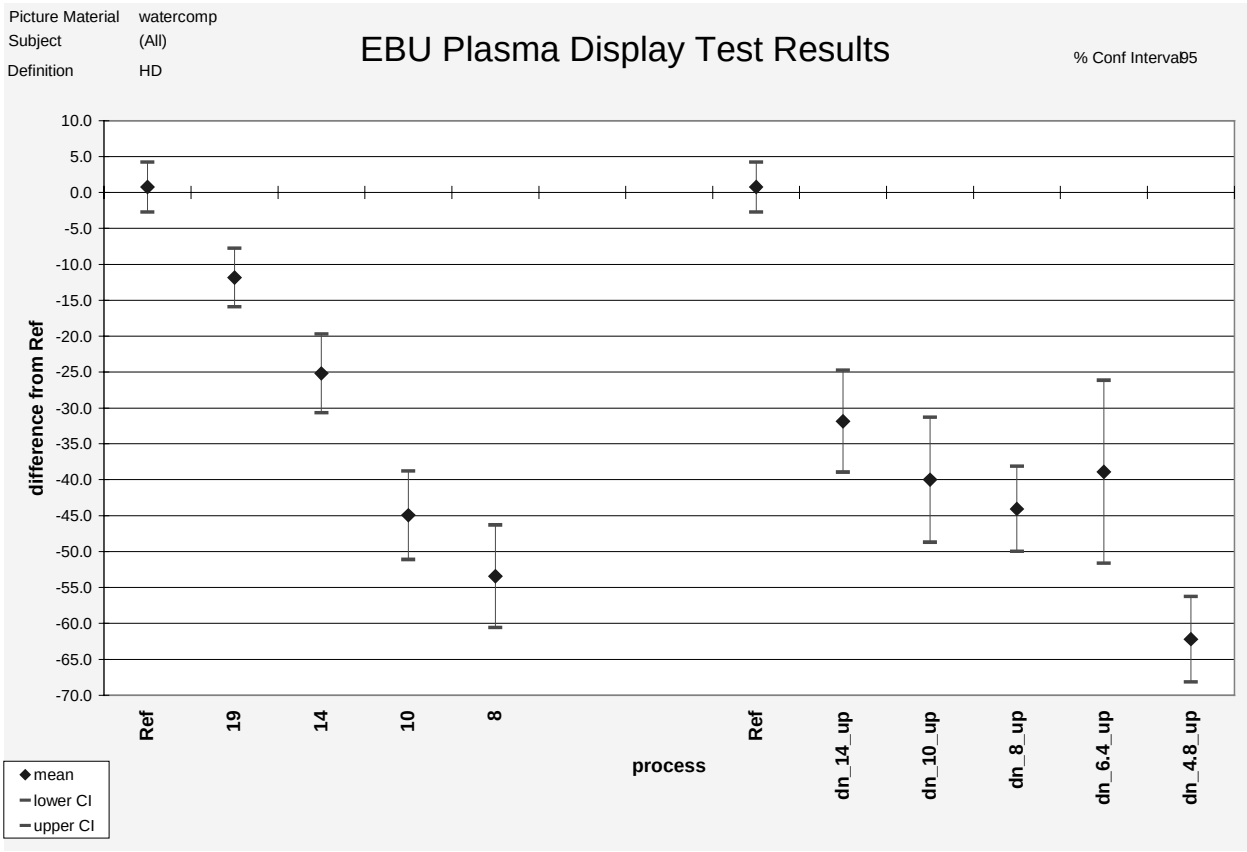

Picture Material panslow
Subject (All)
Definition HD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19
10
8
6.4
Ref
dn_10_up
dn_8_up
dn_6.4_up
dn_4.8_up
process
mean
lower CI
upper CI

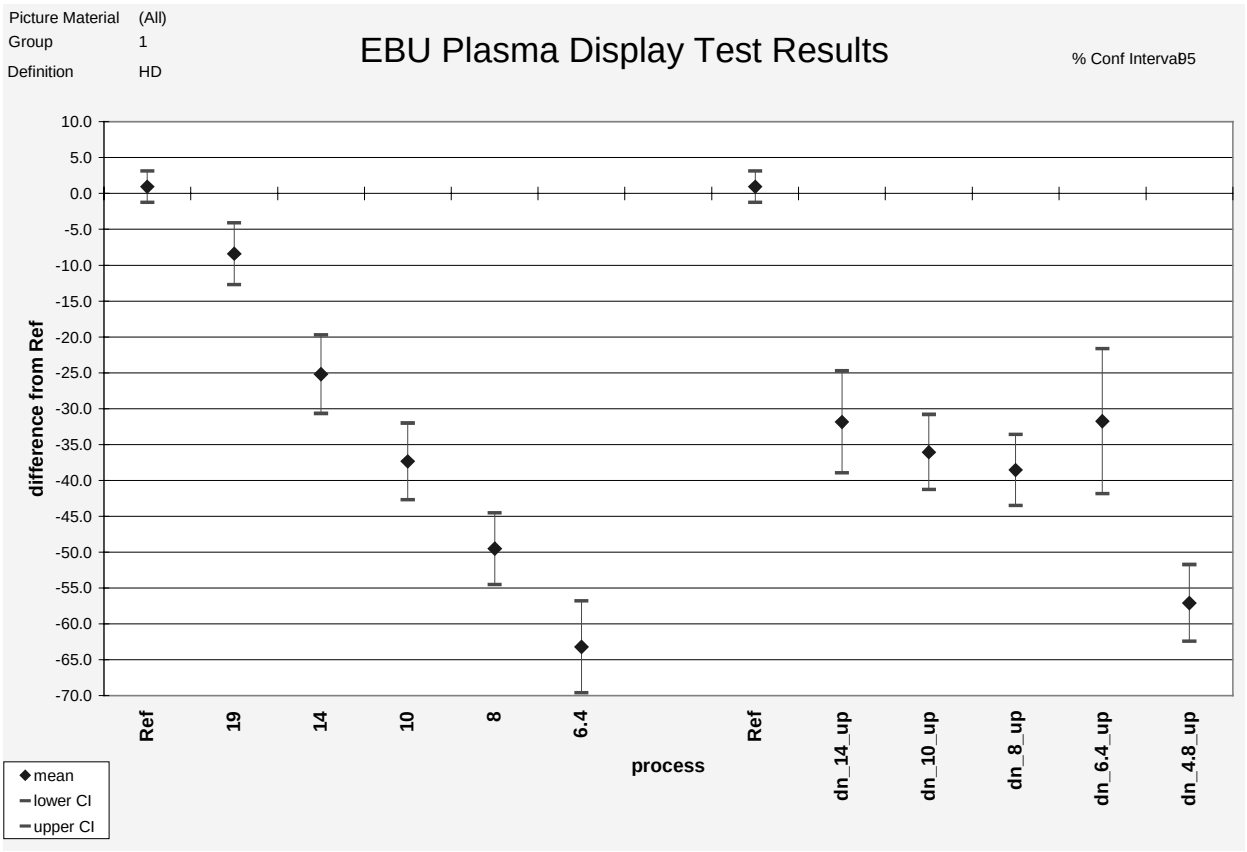

Picture Material waterfall
Subject (All)
Definition HD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19
10
8
6.4
Ref
dn_10_up
dn_8_up
dn_6.4_up
dn_4.8_up
process
mean
lower CI
upper CI

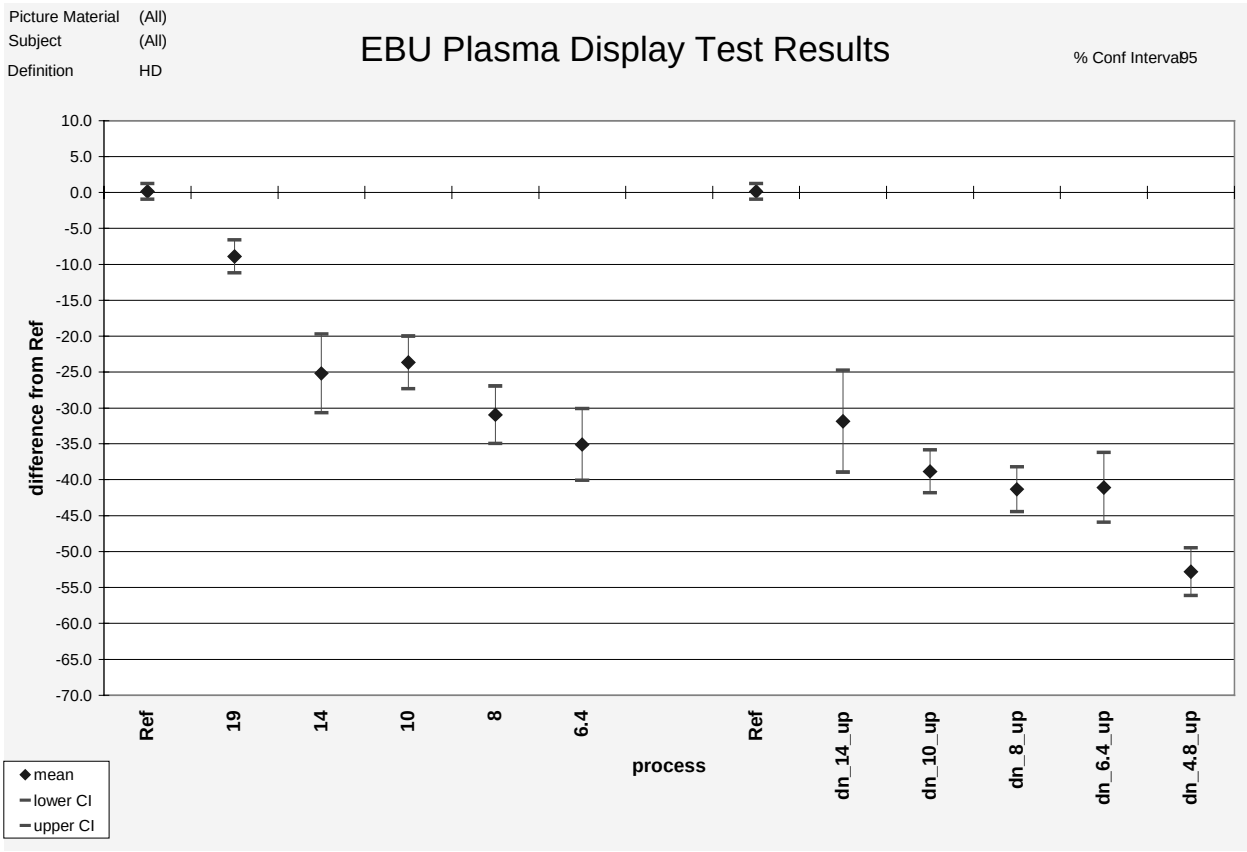

Picture Material tennis
Subject (All)
Definition HD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19
10
8
6.4
Ref
dn_10_up
dn_8_up
dn_6.4_up
dn_4.8_up
process
mean
lower CI
upper CI


Picture Material (All)
Group 0
Definition HD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19
10
8
6.4
Ref
dn_10_up
dn_8_up
dn_6.4_up
dn_4.8_up
process
mean
lower CI
upper CI


Picture Material swim-reflection
Subject (All)
Definition HD
EBU Plasma Display Test Results
% Conf Interval 95
difference from Ref
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
Ref
19
10
8
6.4
Ref
dn_10_up
dn_8_up
dn_6.4_up
dn_4.8_up
process
mean
lower CI
upper CI


Picture Material watercomp
Subject (All)
Definition HD
EBU Plasma Display Test Results
% Conf Interval 95
difference from Ref
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
Ref
19
14
10
8
Ref
dn_14_up
dn_10_up
dn_8_up
dn_6.4_up
dn_4.8_up
process
mean
lower CI
upper CI


Picture Material (All)
Group 1
Definition HD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19
14
10
8
6.4
Ref
dn_14_up
dn_10_up
dn_8_up
dn_6.4_up
dn_4.8_up
process
mean
lower CI
upper CI


Picture Material (All)
Subject (All)
Definition HD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19
14
10
8
6.4
Ref
dn_14_up
dn_10_up
dn_8_up
dn_6.4_up
dn_4.8_up
process
mean
lower CI
upper CI

4.3 480 line PDP results

The results of the SD test, involving the 480 line PDP, are shown in Figs. 11 – 18. As before, these show means and 95% confidence limits of the difference between the score for the reference and the score for each processing condition. The conditions for coding followed by down-conversion (HD delivery) are grouped before the conditions for down-conversion followed by coding (SD delivery). Although the results for the pictures do not vary so widely as in the HD test, they are shown for each picture separately and for the averages, in the same order as before.

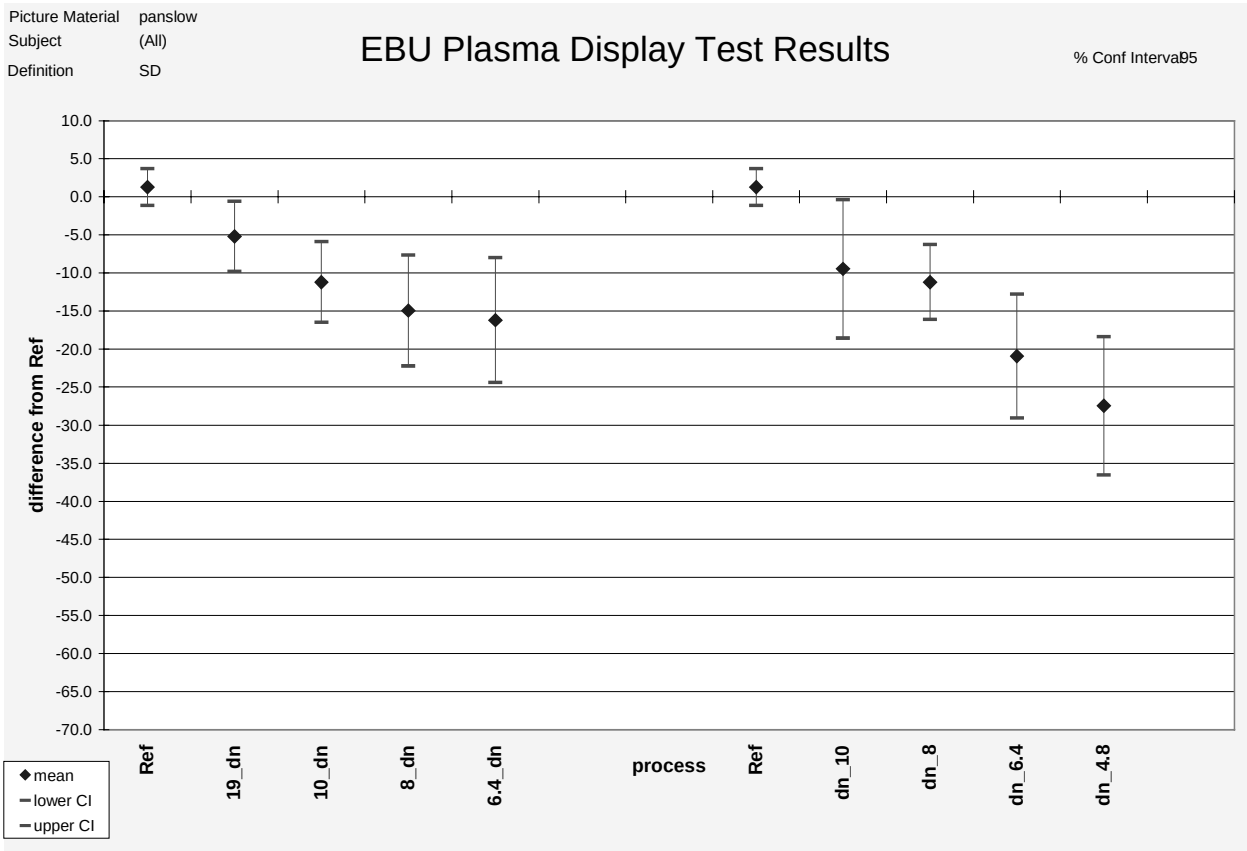
The dominant feature of the results is the significant size of the confidence intervals compared to the overall spread of results in the tests. This arises from the lower resolution and greater viewing distances in this tests, which made the subjects' task much more difficult. As a result, it is hard to make statements regarding particular comparisons that are significant.

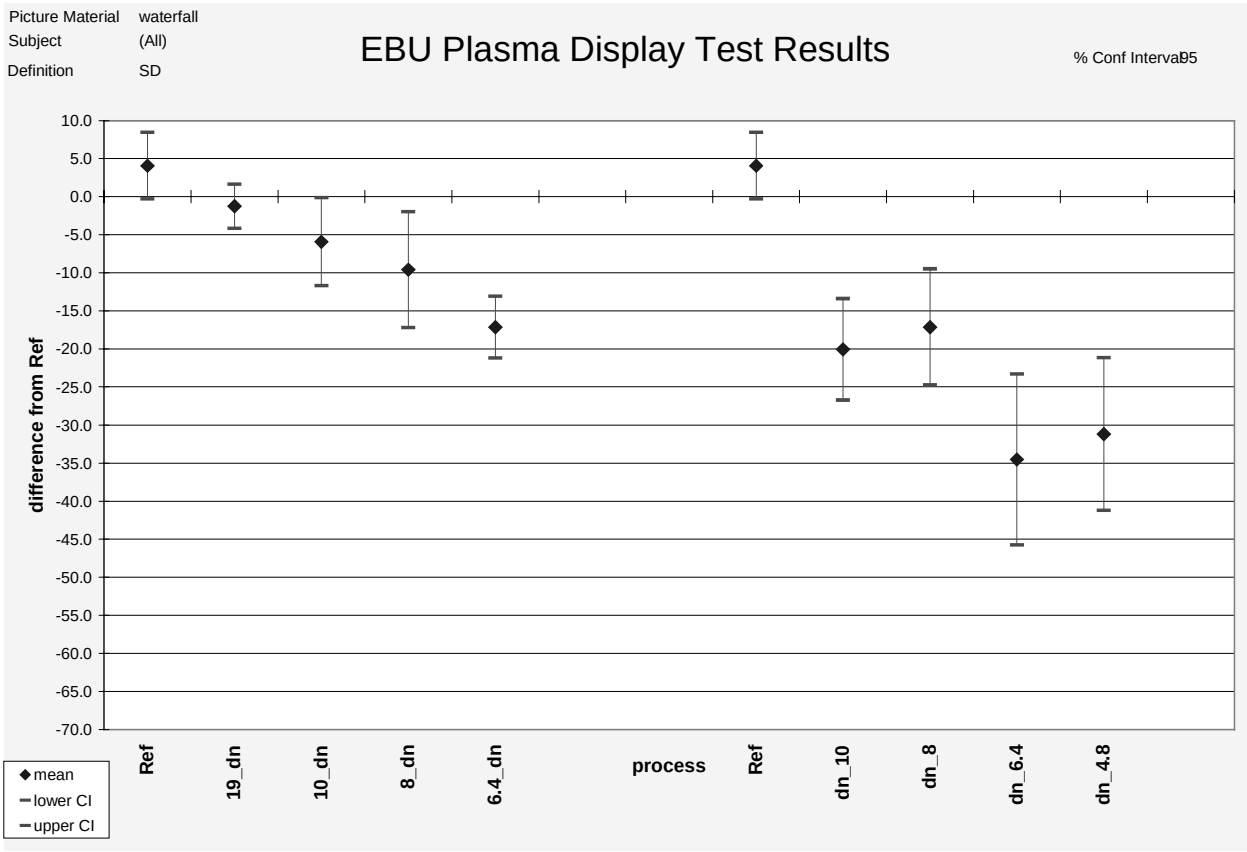
Following the results for the HD test, the pictures have been divided into the same two groups. For the pictures *panslow* and *waterfall* HD delivery is either insignificantly different from, or marginally preferred to, SD delivery. On the other hand, the results for *swim-reflection* and *watercomp* show that, at bit rates at or below 10Mb/s, SD delivery is insignificantly different from, or marginally prefered to, HD delivery.

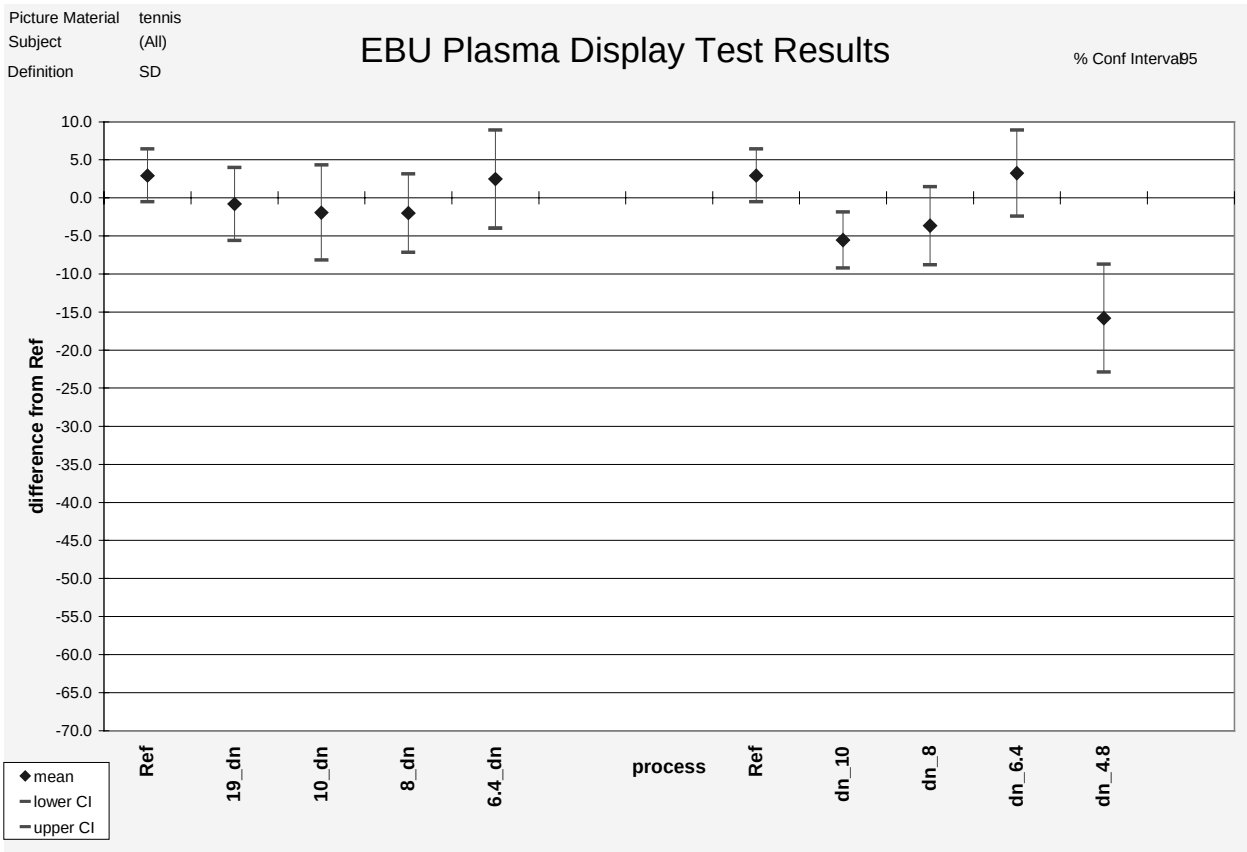
The results for *tennis* show that this picture was not so testing, as the differences from the reference are smaller. There is no significant difference between HD and SD delivery.

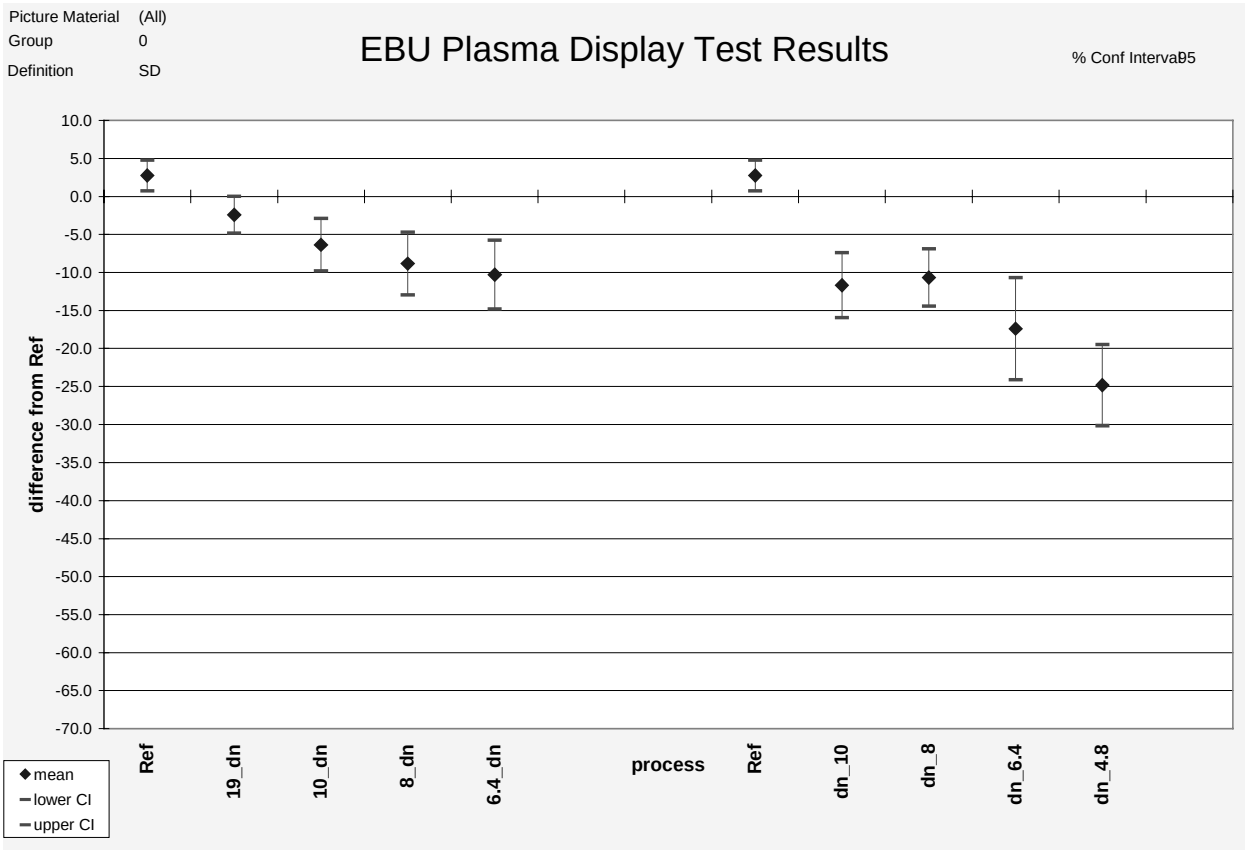
The averages for the groups show this preference a little more clearly, but the average of the groups shows no significant advantage for HD or SD delivery.

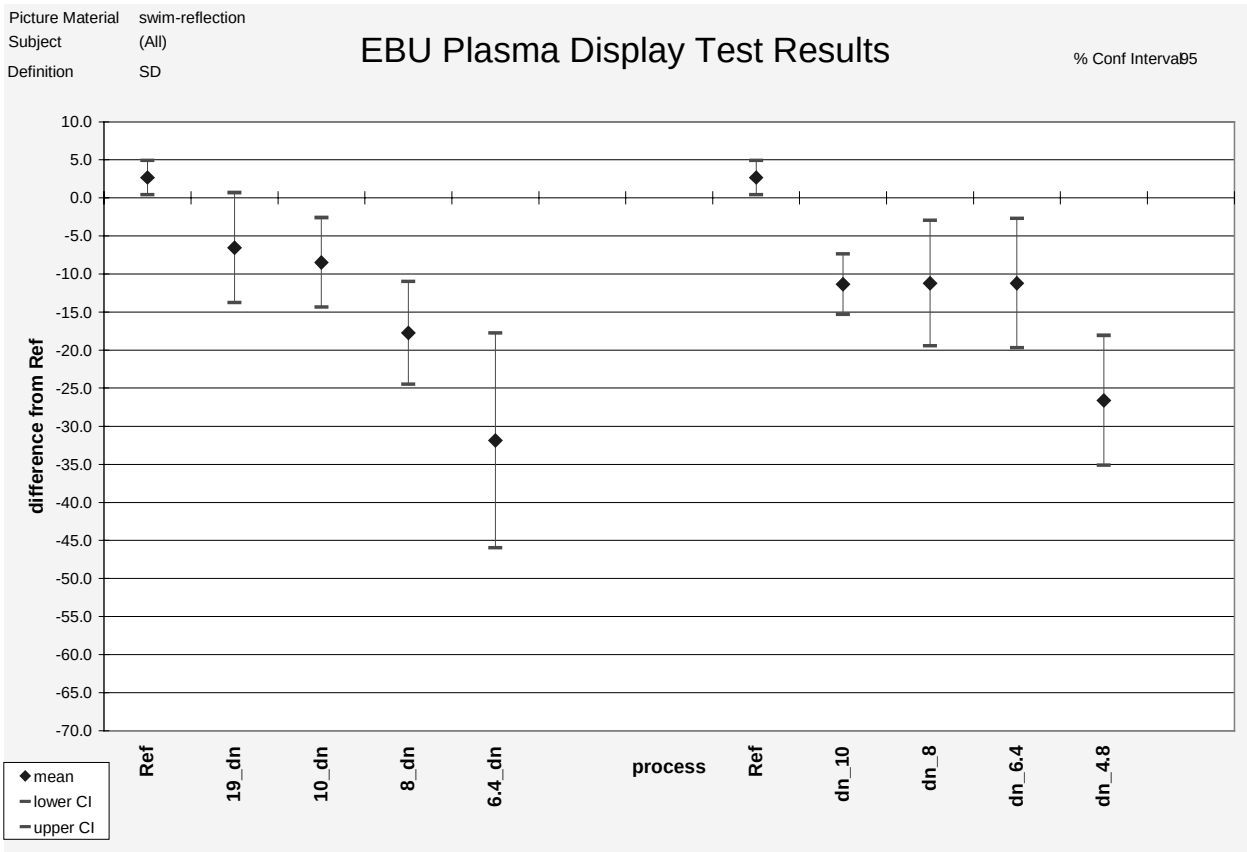
Figs. 11 – 18, following: Results of the SD test for individual pictures and for group averages

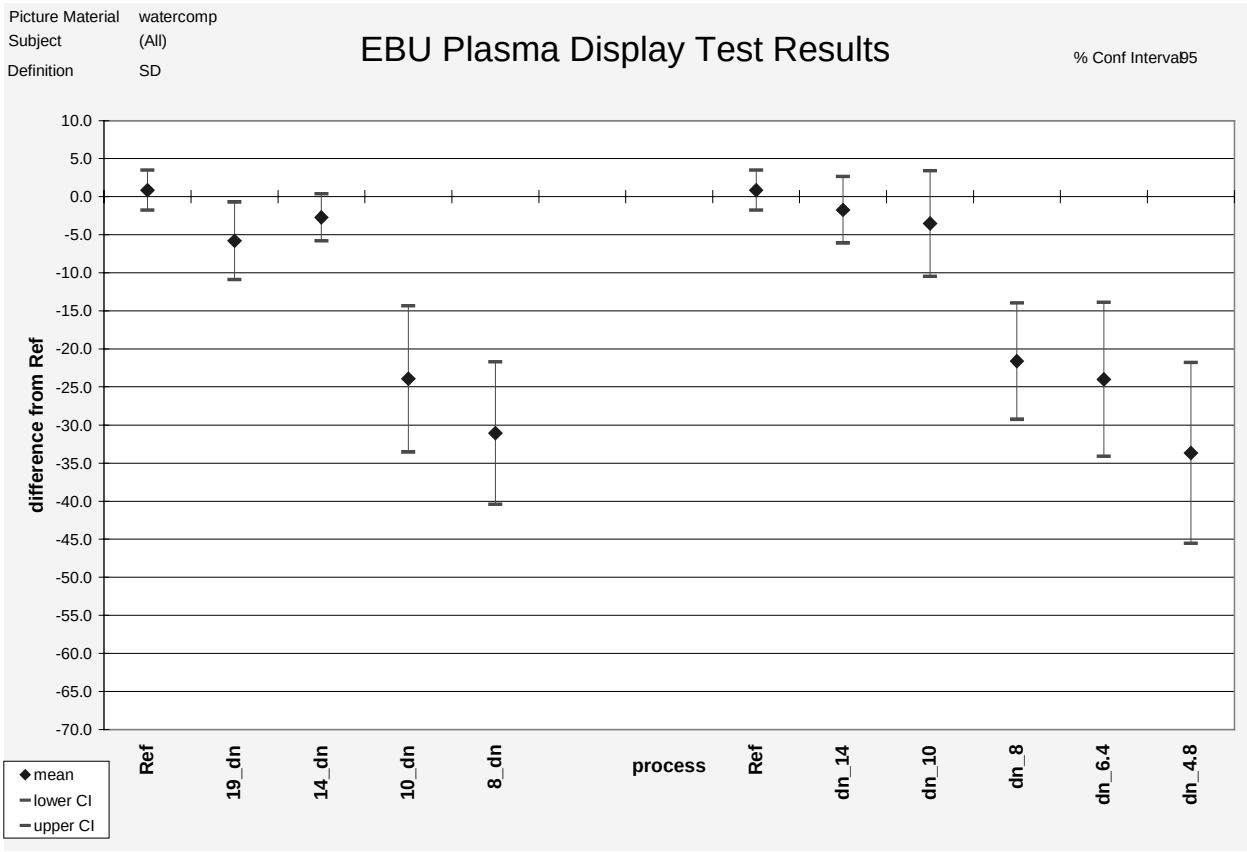

Picture Material panslow
Subject (All)
Definition SD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19_dn
10_dn
8_dn
6.4_dn
process
Ref
dn_10
dn_8
dn_6.4
dn_4.8
mean
lower CI
upper CI

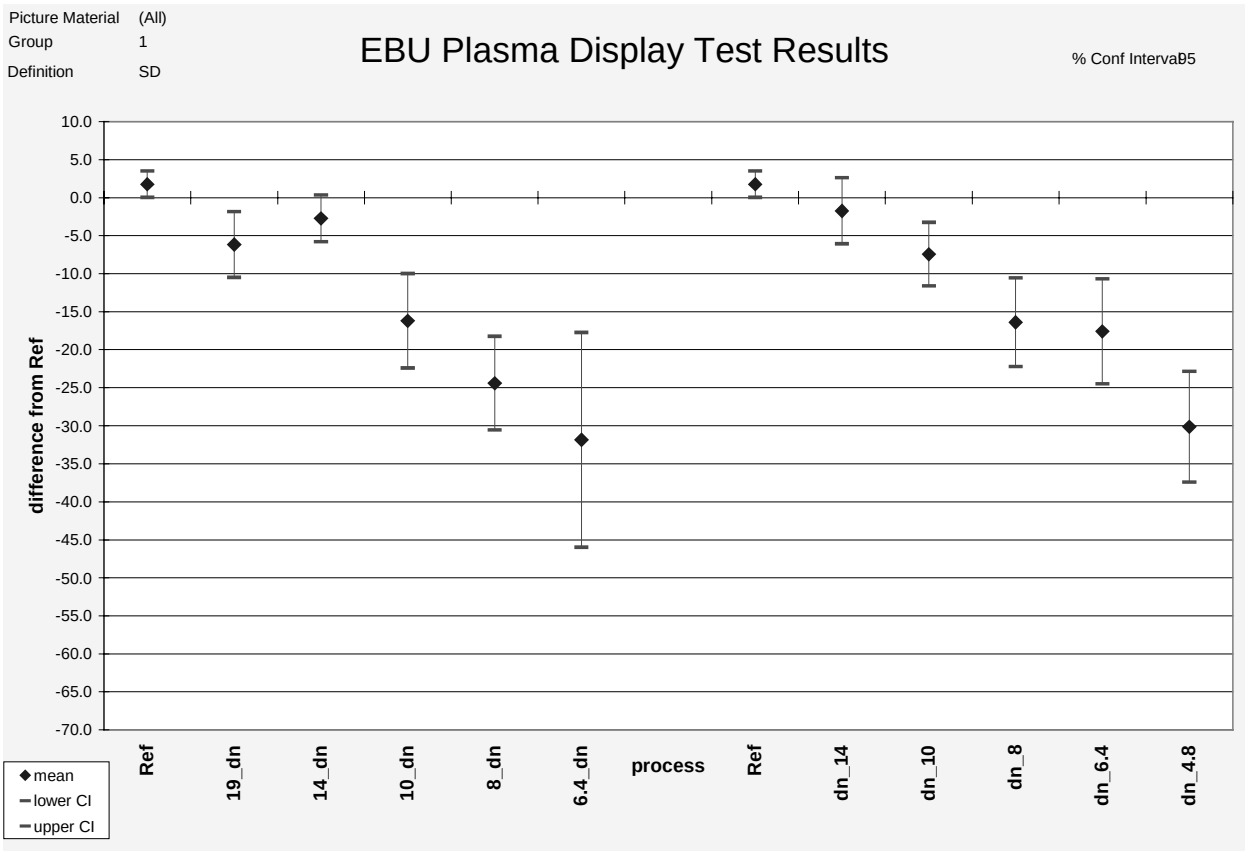

Picture Material waterfall
Subject (All)
Definition SD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19_dn
10_dn
8_dn
6.4_dn
process
Ref
dn_10
dn_8
dn_6.4
dn_4.8
mean
lower CI
upper CI

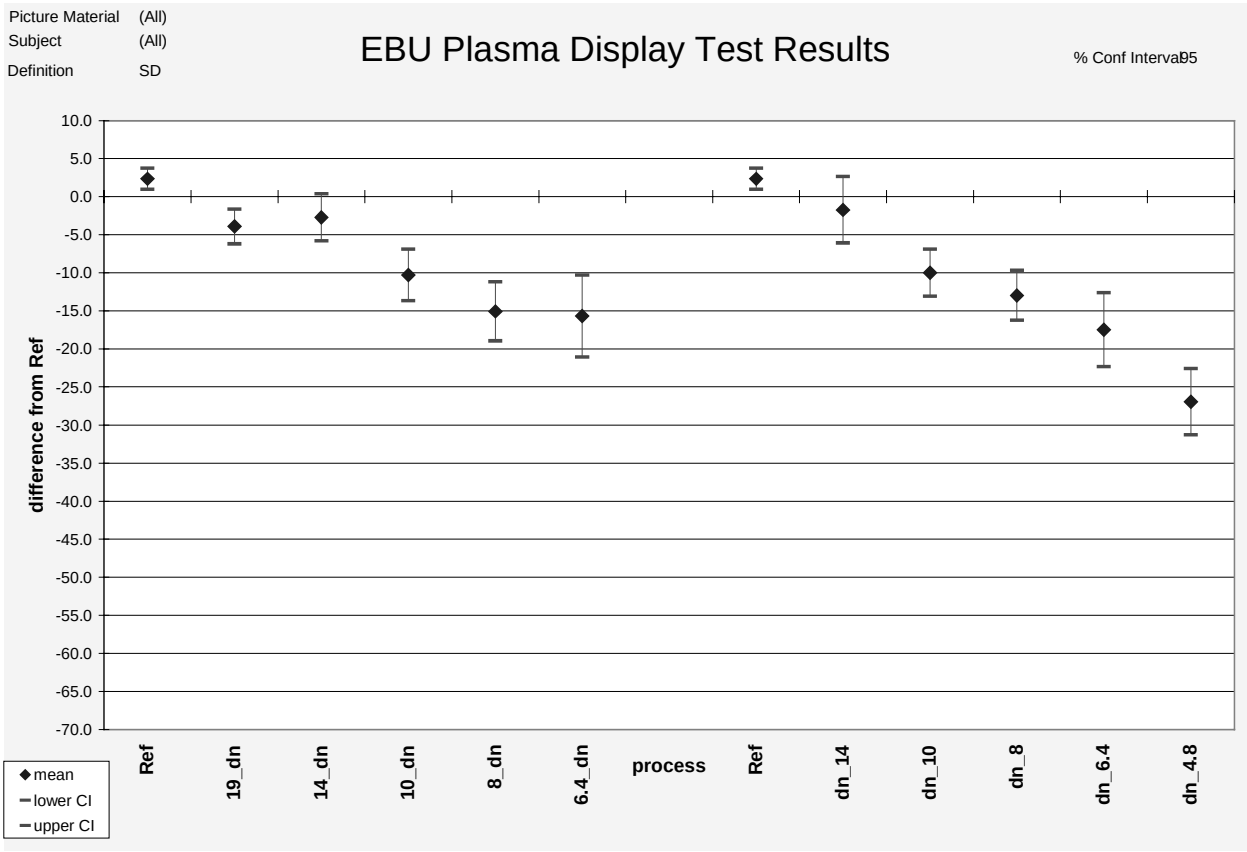

Picture Material tennis
Subject (All)
Definition SD
EBU Plasma Display Test Results
% Conf Interval 95
difference from Ref
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
Ref
19_dn
10_dn
8_dn
6.4_dn
process
Ref
dn_10
dn_8
dn_6.4
dn_4.8
mean
lower CI
upper CI


Picture Material (All)
Group 0
Definition SD
EBU Plasma Display Test Results
% Conf Interval 95
difference from Ref
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
Ref
19_dn
10_dn
8_dn
6.4_dn
process
Ref
dn_10
dn_8
dn_6.4
dn_4.8
mean
lower CI
upper CI


Picture Material swim-reflection
Subject (All)
Definition SD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19_dn
10_dn
8_dn
6.4_dn
process
Ref
dn_10
dn_8
dn_6.4
dn_4.8
mean
lower CI
upper CI


Picture Material watercomp
Subject (All)
Definition SD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19_dn
14_dn
10_dn
8_dn
process
Ref
dn_14
dn_10
dn_8
dn_6.4
dn_4.8
mean
lower CI
upper CI


Picture Material (All)
Group 1
Definition SD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19_dn
14_dn
10_dn
8_dn
6.4_dn
process
Ref
dn_14
dn_10
dn_8
dn_6.4
dn_4.8
mean
lower CI
upper CI


Picture Material (All)
Subject (All)
Definition SD
EBU Plasma Display Test Results
% Conf Interval 95
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
-25.0
-30.0
-35.0
-40.0
-45.0
-50.0
-55.0
-60.0
-65.0
-70.0
difference from Ref
Ref
19_dn
14_dn
10_dn
8_dn
6.4_dn
process
Ref
dn_14
dn_10
dn_8
dn_6.4
dn_4.8
mean
lower CI
upper CI

5 Comments on the test results

Based on the above results the following conclusions can be drawn. For the HD test, we can say that:

- Overall, 19 Mb/s is needed to ensure that pictures on a HD display delivered by HD are significantly better than when delivered by SD;
- At lower rates some pictures were better delivered via SD;

For the SD test, we can say:

- Overall, 8-10 Mb/s is required for delivery to an SD PDP display to produce results within half a grade of the original picture;
- Some picture sequences were slightly better delivered to an SD PDP via coding at HD at bit rates over 6.4 Mb/s.

References

[1] S. Mantovano, M. Stroppiana & P. Sunna, ‘Quality versus bit-rates for Standard Definition TV’ Centro Ricerche e InnovazioneTecnologica, RAI

[2] ITU Recommendation 500-4, ‘Method for the subjective assessment of the quality of television pictures’

[3] MPEG Software Simulation Group http://www.mpeg.org/MPEG/MSSG

Appendix A: Instructions to viewers

The following were the instructions read out to participants in the tests.

Background

The tests you are about to take part in are being performed in conjunction with the EBU Displays Group. The aim of the tests is to investigate the effects of MPEG-2 compression on pictures displayed on large flat panel displays and to relate bit rates, scanning formats and display technology to subjective quality.

Specifically, the tests will compare the delivery of High Definition TV with Standard Definition, at a variety of bit rates, to viewers equipped with large displays.

Description of the test

In the test you will view a number of comparisons between two video sequences. Each comparison will last a minute. You will see a caption giving the number of the test, followed by the first sequence – Sequence A – then the second sequence – Sequence B. As you watch the pictures the light box above the screen will illuminate to tell you which is which. There will then be a repetition of the comparison. Finally there will be a scoring period of about 10 seconds before the next test.

Your job is to assign a mark to both of the two sequences in the scoring period. Please only mark your sheet during the scoring period, not before. If you look at your score-sheet you will see that for each test number you have two 10cm lines on which you can mark a point for both sequence A and sequence B. This is a **continuous** scale, and the wording on the side (‘excellent’, ‘fair’ etc) is there only for guidance. In particular, a mark at the higher end of ‘good’ will count for more than one at the lower end.

If you change your mind or make a mistake, cross out your first score clearly and make another mark.

Things to look for

You will see two main types of impairment with the picture that might affect your scoring: compression impairments, and conversion impairments which result from displaying on the flat panel display.

Compression impairments might include blockiness, fuzziness, noise in localised regions of the picture or noise which seems to come and go over time. You might also see impairments near sharp edges.

Conversion impairments might also include patterning on the picture, such as jagged edges, which will often move in a different way to the picture.

When judging the pictures please take all these factors into account as well as any other factors which you judge to be important in the quality of the picture.

Points to bear in mind

- You must score both sequence A and sequence B;
- The score line is continuous;

- Do not assume anything about the order in which the sequences are presented to you: in particular how each sequence was encoded or compressed;
- You might not find it easy to tell the difference between sequences.

Questions

Any questions?

Demonstration

[Consisting of two test conditions derived from *spincalendar1*]

Appendix B: SSG MPEG-2 codec parameter files

B.1 576i parameter file

```
MPEG-2 Test Sequence, 25 frames/sec
/pics/thomasd/HDTV/pic_files/EBU_pics/576I/panslow/panslow_dn576i  /* name of
source files */
-         /* name of reconstructed images (-: don't store) */
-         /* name of intra quant matrix file     (-: default matrix) */
inter.mat /* name of non intra quant matrix file (-: default matrix) */
-    /* name of statistics file (-: stdout ) */
4         /* input picture file format: 0=*.Y,*.U,*.V, 1=*.yuv, 2=*.ppm 3=*.SIF
4= KW */
300      /* number of frames */
0         /* number of first frame */
00:00:00:00 /* timecode of first frame */
12        /* N (# of frames in GOP) */
3         /* M (I/P frame distance) */
0         /* ISO/IEC 11172-2 stream */
0         /* 0:frame pictures, 1:field pictures */
720       /* horizontal_size */
576       /* vertical_size */
3         /* aspect_ratio_information 1=square pel, 2=4:3, 3=16:9, 4=2.11:1 */
3         /* frame_rate_code 2=23.976, 3=25, 4=29.97, 5=30 frames/second */
10000000 /* bit_rate (bits/s) */
112       /* vbv_buffer_size (in multiples of 16 kbit) */
0         /* low_delay  */
0         /* constrained_parameters_flag */
4         /* Profile ID: Simple = 5, Main = 4, SNR = 3, Spatial = 2, High = 1 */
4         /* Level ID:   Low = 10, Main = 8, High 1440 = 6, High = 4          */
0         /* progressive_sequence */
1         /* chroma_format: 1=4:2:0, 2=4:2:2, 3=4:4:4 */
0         /* video_format: 0=comp., 1=PAL, 2=NTSC, 3=SECAM, 4=MAC, 5=unspec. */
5         /* color_primaries */
5         /* transfer_characteristics */
5         /* matrix_coefficients */
720       /* display_horizontal_size */
576       /* display_vertical_size */
0         /* intra_dc_precision (0: 8 bit, 1: 9 bit, 2: 10 bit, 3: 11 bit */
1         /* top_field_first */
0 0 0     /* frame_pred_frame_dct (I P B) */
0 0 0     /* concealment_motion_vectors (I P B) */
1 1 1     /* q_scale_type  (I P B) */
1 0 0     /* intra_vlc_format (I P B)*/
1 1 1     /* alternate_scan (I P B) */
0         /* repeat_first_field */
0         /* progressive_frame */
0         /* P distance between complete intra slice refresh */
0         /* rate control: r (reaction parameter) */
0         /* rate control: avg_act (initial average activity) */
0         /* rate control: Xi (initial I frame global complexity measure) */
0         /* rate control: Xp (initial P frame global complexity measure) */
0         /* rate control: Xb (initial B frame global complexity measure) */
0         /* rate control: d0i (initial I frame virtual buffer fullness) */
0         /* rate control: d0p (initial P frame virtual buffer fullness) */
0         /* rate control: d0b (initial B frame virtual buffer fullness) */
4 4 54 39 /* P:  forw_hor_f_code forw_vert_f_code search_width/height */
3 2 18 13 /* B1: forw_hor_f_code forw_vert_f_code search_width/height */
4 3 36 26 /* B1: back_hor_f_code back_vert_f_code search_width/height */
4 3 36 26 /* B2: forw_hor_f_code forw_vert_f_code search_width/height */
3 2 18 13 /* B2: back_hor_f_code back_vert_f_code search_width/height */
```

B.2 720p parameter file

```
MPEG-2 Test Sequence, 50 frames/sec progressive
/pics/thomasd/HDTV/pic_files/EBU_pics/panslow/panslow  /* name of source files
*/
-         /* name of reconstructed images (-: don't store) */
-         /* name of intra quant matrix file     (-: default matrix) */
inter.mat /* name of non intra quant matrix file (-: default matrix) */
-    /* name of statistics file (-: stdout ) */
4         /* input picture file format: 0=*.Y,*.U,*.V, 1=*.yuv, 2=*.ppm 3=*.SIF
4= KW */
600      /* number of frames */
0         /* number of first frame */
00:00:00:00 /* timecode of first frame */
24        /* N (# of frames in GOP) */
6         /* M (I/P frame distance) */
0         /* ISO/IEC 11172-2 stream */
0         /* 0:frame pictures, 1:field pictures */
1280       /* horizontal_size */
720       /* vertical_size */
3         /* aspect_ratio_information 1=square pel, 2=4:3, 3=16:9, 4=2.11:1 */
6         /* frame_rate_code 2=23.976, 3=25, 4=29.97, 5=30 frames/second */
19000000 /* bit_rate (bits/s) */
597       /* vbv_buffer_size (in multiples of 16 kbit) */
0         /* low_delay  */
0         /* constrained_parameters_flag */
4         /* Profile ID: Simple = 5, Main = 4, SNR = 3, Spatial = 2, High = 1 */
4         /* Level ID:   Low = 10, Main = 8, High 1440 = 6, High = 4          */
1         /* progressive_sequence */
1         /* chroma_format: 1=4:2:0, 2=4:2:2, 3=4:4:4 */
0         /* video_format: 0=comp., 1=PAL, 2=NTSC, 3=SECAM, 4=MAC, 5=unspec. */
5         /* color_primaries */
5         /* transfer_characteristics */
5         /* matrix_coefficients */
1280       /* display_horizontal_size */
720       /* display_vertical_size */
0         /* intra_dc_precision (0: 8 bit, 1: 9 bit, 2: 10 bit, 3: 11 bit */
1         /* top_field_first */
0 0 0     /* frame_pred_frame_dct (I P B) */
0 0 0     /* concealment_motion_vectors (I P B) */
1 1 1     /* q_scale_type  (I P B) */
1 0 0     /* intra_vlc_format (I P B)*/
0 0 0     /* alternate_scan (I P B) */
0         /* repeat_first_field */
1         /* progressive_frame */
0         /* P distance between complete intra slice refresh */
0         /* rate control: r (reaction parameter) */
0         /* rate control: avg_act (initial average activity) */
0         /* rate control: Xi (initial I frame global complexity measure) */
0         /* rate control: Xp (initial P frame global complexity measure) */
0         /* rate control: Xb (initial B frame global complexity measure) */
0         /* rate control: d0i (initial I frame virtual buffer fullness) */
0         /* rate control: d0p (initial P frame virtual buffer fullness) */
0         /* rate control: d0b (initial B frame virtual buffer fullness) */
5 4 96 48 /* P:  forw_hor_f_code forw_vert_f_code search_width/height */
3 2 16 8 /* B1: forw_hor_f_code forw_vert_f_code search_width/height */
5 4 80 40 /* B1: back_hor_f_code back_vert_f_code search_width/height */
4 3 32 16 /* B2: forw_hor_f_code forw_vert_f_code search_width/height */
5 4 64 32 /* B2: back_hor_f_code back_vert_f_code search_width/height */
4 3 48 24 /* B3: forw_hor_f_code forw_vert_f_code search_width/height */
4 3 48 24 /* B3: back_hor_f_code back_vert_f_code search_width/height */
5 4 64 32 /* B4: forw_hor_f_code forw_vert_f_code search_width/height */
4 3 32 16 /* B4: back_hor_f_code back_vert_f_code search_width/height */
```

```
5 4 80 40 /* B5: forw_hor_f_code forw_vert_f_code search_width/height */
3 2 16 8 /* B5: back_hor_f_code back_vert_f_code search_width/height */
```

B.3 Non-intra block quantization matrix

```
16 17 18 19 20 21 22 23
17 18 19 20 21 22 23 24
18 19 20 21 22 23 24 25
19 20 21 22 23 24 25 26
20 21 22 23 24 25 26 27
21 22 23 24 25 26 27 28
22 23 24 25 26 27 28 29
23 24 25 26 27 28 29 30
```