Watermarking – summary results of EBU tests

Louis Cheveau *(EBU)*, **Eddy Goray** *(RTBF)*
and Richard Salmon *(BBC R&D)*

EBU members – within the framework of the N/WTM project group – recently carried out tests on four watermarking systems, from a subjective, objective and robustness viewpoint. This article offers a brief summary of the results of these tests.

Two previous articles in **EBU Technical Review** [1][2] discussed the need for transmission security on digital networks such as the EBU's *Eurovision* network. The technical details of the proposed security system were described in [2].

Following on from these earlier articles, the present article reports on the results of tests carried out on four watermarking systems – supplied by Lucent, Philips, Tektronix and Thomson. Subjective evaluations on these systems were carried out by the RAI, IRT and SVT, while some objective assessments were conducted by TDF. Furthermore, an extensive series of robustness tests was carried out by the BBC.

System parameters

Only the Philips and Thomson systems complied with:

1) the 64-bit data-capacity-per-WMS (Watermark Minimum Segment) requirement;
2) the data interface that was specified to assist with the testing procedure.

The Lucent and Tektronix systems had to be adapted, to some extent, to enable the testing.

The Philips and Thomson systems also provided the capability of using a large number of orthogonal watermark channels. The Lucent system could provide a very limited number of orthogonal channels – the data capacity of a channel being 24 bits in a 1-second WMS, or 120 bits in a 5-second WMS.

The Tektronix system – which had been designed for a rather different application (audio/video synchronization) – was modified by the manufacturer for the purpose of these tests. In this form, it operated at the same level of perceptibility but with a reduced data channel of about 21 bits in a 5-second WMS. The data capacity in unmodified units, although not tested by the EBU, is higher and provides for both an identification channel and a representation of the audio envelope.

Only the Philips and Thomson systems were designed to be used with a secret key and, furthermore, these were the only manufacturers who confirmed that they are willing to release the algorithm under NDA (non-disclosure agreement) for evaluation by an independent expert.

The differences in data capacity made a comparison between the Philips and Thomson systems – and the other two systems – almost impossible since any watermarking system is a compromise between:

- the payload capacity;
- the level of visibility/invisibility of the watermark, i.e., the "strength" of the embedding;
- the robustness to possible attacks.

Test results

There were three categories of test results: *subjective*, *objective* and *robustness*.

Abbreviations

BBC	British Broadcasting Corporation
CRT	Cathode ray tube
DVE	Digital video-effects unit
DSCQS	Double-stimulus continuous-quality scale
IRT	*Institut für Rundfunktechnik GmbH* (German broadcast engineering research centre)
ITU	International Telecommunication Union
ITU-R	ITU - Radiocommunication Sector
JPEG	Joint Photographic Experts Group
MP@ML	(MPEG-2) Main Profile at Main Level
MPEG	Moving Picture Experts Group
N/A	Not applicable
NDA	Non-disclosure agreement
PAL	Phase alternation line
RAI	*Radiotelevisione Italiana*
RGB	Red-green-blue colour model
SDI	Serial digital interface
SVT	*Sveriges Television och Radio Grupp* (Sweden)
TBD	To be decided
WMS	Watermark minimum segment
YUV	The luminance (Y) and colour difference (U and V) signals of the PAL colour television system

Subjective quality tests

General results

Results in this category are available from RAI, the IRT and SVT.

a) RAI (Fig. 1) and IRT (Fig. 2)

The tests carried out by RAI and the IRT were performed in accordance with *Paragraph 5*, *Variant II* of ITU-R Recommendation BT.500 [3] — the double-stimulus, continuous-quality-scale (DSCQS) method. The vertical scale on the graphs shown in *Figs. 1 and 2* represent the difference in perceived quality between the test items and the source.

b) SVT (Fig. 3)

The tests carried out by SVT were performed as a comparison test, where the assessors were told to judge which of the two scenes, A or B, had been subject to watermarking. The following grading scale was used:

A much worse than B	(Corresponds to 100%)
A worse than B	
A slightly worse then B	
A equal to B	(Corresponds to 0%)
B slightly worse than A	
B worse than A	
B much worse than A	(Corresponds to 100%)

The observers were told to make their judgement on a continuous scale which had been provided on the test forms.

The scores are depicted graphically in *Fig. 3*. A score corresponding to 33% means "slightly worse than", 67% "worse than" and 100% "much worse than" the reference.

Apart from a few exceptions (Belgian landscape), the results from RAI, the IRT and SVT suggest that the picture quality would be adequate for current domestic viewing conditions. The RAI and IRT tests considered watermarking in the production environment, and also in a contribution environment, with an additional watermark added after a contribution link. The SVT tests only considered the production environment.

Figure 1
The RAI test results.

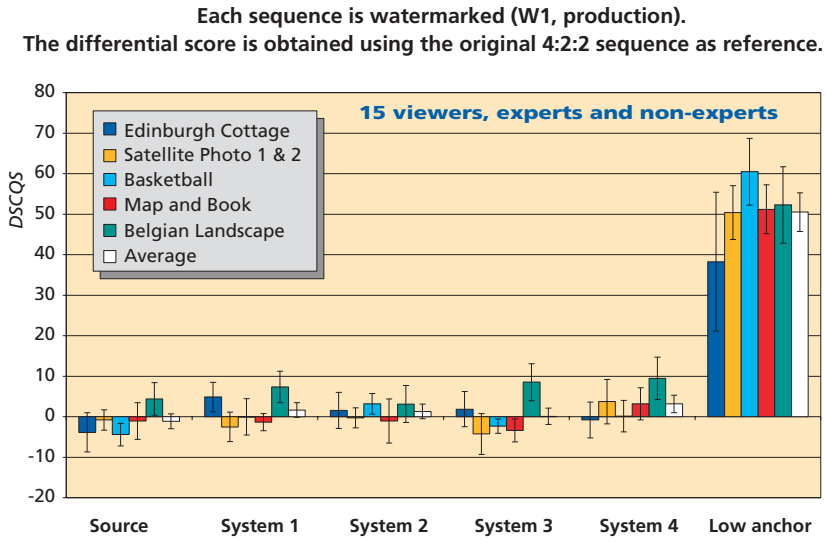


Figure 2
The IRT test results.

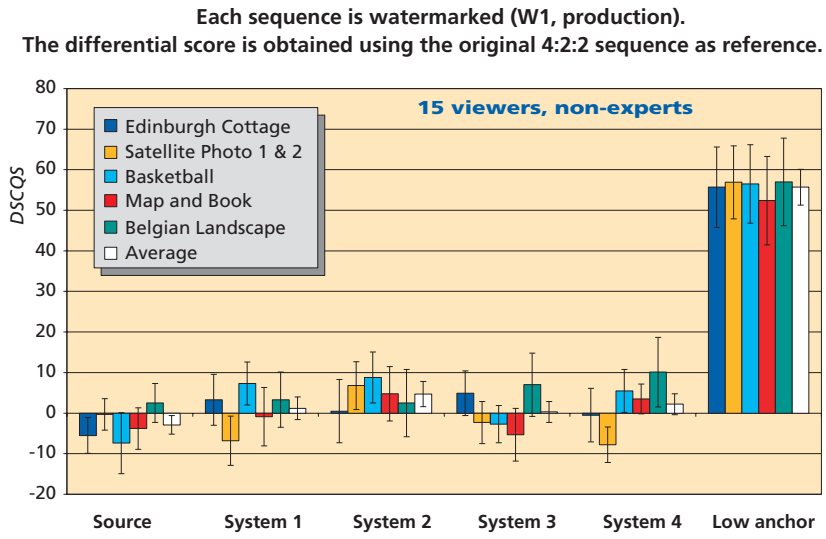
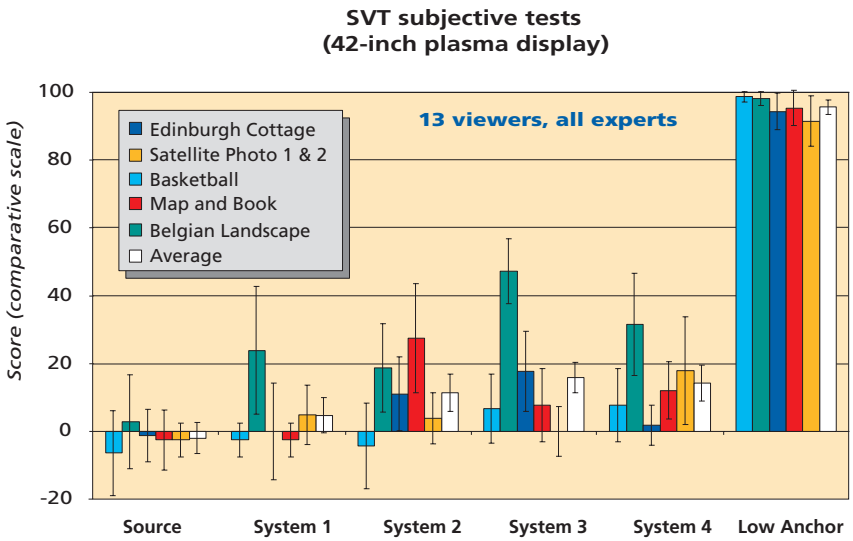


Figure 3
The SVT test results.



The results shown in *Figs. 1 - 3* are limited to the production environment only (i.e. no MPEG compression was used here).

Expert viewing

Whilst the subjective test results are good in a general sense, it should be noted that, during the tests, some expert viewing was made at the BBC and that the experts were definitely able – at least in some cases – to recognize the presence of watermarking.

Both flat-panel and CRT displays were used during the tests. At the BBC, when using a plasma display with a direct digital input, it was found that the watermark was slightly more visible than in the CRT case. At RAI, the conclusion was that the use of flat panel displays (with analogue inputs) did not change the results of the subjective assessment tests made with CRT displays.

Whilst the subjective test results indicate that the watermarking systems might be largely invisible to non-expert viewers, if watermarks are to be used in a production environment, more critical expert viewing is required. In this environment, a lower strength of watermark embedding may be appropriate. To maintain the robustness, we may have to relax our data payload requirement.

Cascading

In all cases, it was found that cascading a watermark with itself caused a substantial increase in its visibility.

Tests which involved the cascading of the various watermark systems with each other indicated that the systems were orthogonal, in that they did not appear to interfere with each other. Also, there was no noticeable increase in watermark visibility when different systems were cascaded.

Objective results

Objective quality measurements were provided by TDF, based on the MAEVA equipment. In order to fully "train" this equipment, more test material would have been required. Nevertheless, the preliminary objective results were broadly in line with the subjective tests, indicating not only the practical invisibility of most watermarked sequences, but also that "Belgian landscape" was by far the most critical sequence.

Robustness tests

The robustness tests were undertaken by BBC R&D (Kingswood Warren). The methodology of these tests is outlined in *Appendix A*.

Several of these tests (those numbered 2 to 12 in *Table A1* of *Appendix A*) made use of a DVE at the BBC premises. This equipment applies some vertical/temporal filtering which appeared to remove the watermark to some extent in three of the systems provided, even with limited scaling or picture shifts. Essentially, only the Philips watermark could to be detected when processed by the DVE. In the scaling tests, it could be detected only up to 140% scaling and down to 70%, in the implementation supplied for the tests. The tests involving picture shifts were also conducted using the DVE, and again the Philips system performed better.

Tektronix was the only system substantially robust to VHS recording, when cascaded with MPEG compression.

As regards MPEG-2 compression (between 2 and 6 Mbit/s), the Philips and Tektronix systems showed the best performance, as shown in *Fig. 4*. This indicates the robustness of the watermarks after contribution coding at 20.5 Mbit/s and subsequent emission coding.

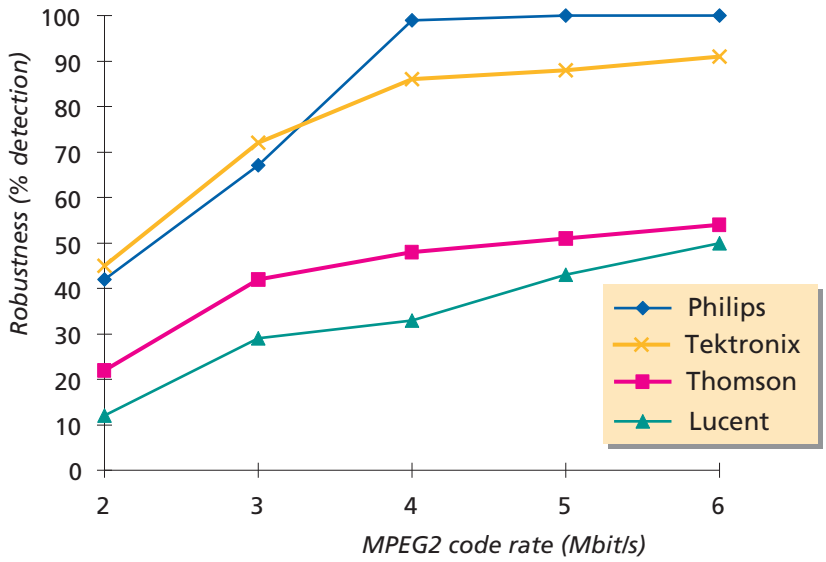


Figure 4
Robustness to emission coding (after contribution coding).

Results obtained from a test that simulated a real digital broadcasting chain were entirely in line with those obtained using the EBU test chain.

Only the Philips and Lucent systems successfully passed the white noise attack.

The slow motion attack was resisted by the Thomson and Philips systems only.

All the systems survived the common professional recording formats.

Conclusions drawn from the tests

In conclusion, the four systems have been optimized with the intention of being invisible. The subjective tests (using the DSCQS method of ITU-R Recommendation BT.500 [3]) showed that, for most of the test signals, the addition of the watermarks did not significantly reduce the picture quality. This does not imply invisibility to expert viewers in studio conditions.

The compromise employed to reach maximum capacity and robustness at the given level of invisibility has been achieved differently by the four manufacturers. Whilst none of the systems completely satisfied the robustness requirements, it is probable that further development, and a reduction in the data capacity requirement from 64 to

Louis Cheveau qualified as a Physics Engineer from the University of Liège, Belgium, in 1967 and obtained a Ph.D. in Physics from the University of Montreal, Canada, in 1974. That year, he joined the EBU Technical Centre in Brussels as head of the computing department and, initially, worked in the field of terrestrial television broadcasting. In 1977, the emphasis of his work changed to satellite broadcasting.

In 1984, Dr Cheveau was detached for two years to CBC in Canada. There, he worked in International Relations with a special emphasis on satellite broadcasting and HDTV matters. In 1986, he returned to the EBU Technical Centre, this time to work on Eurovision transmissions. Since 1989, he has been Head of Transmission Technologies within the EBU Technical Department in Geneva.

Eddy Goray studied Electronic Engineering in Brussels and joined RTBF in 1973 as a Laboratory Engineer. He was head of the R&D Department between 1981 and 1994 and, since 1995, has been head of the TV Diffusion Engineering Department. His main technical background is in digital television, data transmission and TV scrambling systems.

Mr Goray is involved in different projects within the EBU, where he is a member of the Network Technology Management Committee (NMC) and also chairman of two working groups – *Archivex* (Automatic archive exchanges on satellite networks) and *WTM TF* (the EBU Watermarking task force). He has participated actively in different European research projects including OKAPI, Talisman and OCTALIS.

Richard Salmon has worked as a Research Engineer for the BBC at Kingswood Warren, since graduating from the University of Cambridge, UK, in 1987. He has been involved in many different projects over the years, with a particular interest in colorimetry, HDTV and video systems engineering. He also spent six months on attachment to NHK in Japan, working on Plasma Display technology.

For the last two years Mr Salmon has been part of a small team in the BBC, working on watermarking techniques and their application to broadcasting.

maybe 48 bits would improve performance. For information, the original user requirements are summarized in *Table 1*.

Table 1
EBU video-watermarking user requirements for production (W1) and contribution (W2).

Requirements: ✔ = Mandatory, ∇ = Recommended, TBD = To be decided	
1. Visibility Of The Watermark	
⇨ Not visible in a left/right comparison with the original, under studio viewing conditions	✔
2. Payload	
⇨ Watermark Minimum Segment (WMS), duration in time	1 sec for W1, 5 sec for W2
⇨ Data capacity	64 bits / WMS
⇨ Detection probability per WMS	> 95%
⇨ False positive probability per WMS	$< 10^{-8}$
⇨ Probability for (bit) error-free payload per WMS	$> 1 - 10^{-8}$
3. Purpose of the watermark	Identification
4. Security – secret watermarking key	
⇨ Difficult to predict, crytographically strong	✔
⇨ Number of available watermarking keys	A large number (TBD)
⇨ Watermarking-key management	✔
5. Single-ended watermark detection and Payload Extraction	✔
6. Format of Original and Watermarked signals	ITU-T Rec. BT.656
7. Robustness	
7.1 Data compression	
⇨ MJPEG (20 Mbit/s)	✔
⇨ ISO/MPEG-1 (< 1 Mbit/s)	∇
⇨ ISO/MPEG-2 (2 to 6 Mbit/s MP@ML)	✔
⇨ Panasonic/DV or JVC/Digital-S, Sony/DV, Sony/Beta-SX	✔
⇨ MPEG-2 4:2:2 - 50 Mbit/s recorder (Sony IMX)	✔
7.2 PAL-coding and analogue recording	
⇨ Sony/Beta-SP (with PAL input)	✔
⇨ VHS	✔
7.3 Digital and analogue filtering	
⇨ Re-sampling, e.g. D/A → A/D conversion	✔
⇨ Sampling-rate conversion, up and down conversion	✔
⇨ Picture aspect-ratio conversion, e.g. 4:3 ↔ 16:9	✔
⇨ Frame-rate conversion, e.g. 24 Hz ↔ 25 Hz ↔ 30 Hz	✔
⇨ Line-scan conversion: progressive ↔ interlace	✔
⇨ Motion-compensating noise reduction	✔
⇨ Added white noise (at –30 dB)	✔
⇨ Colour-space conversion: colour ↔ grey-scale	✔
⇨ Slow motion 3:1	∇

Table 1
EBU video-watermarking user requirements for production (W1) and contribution (W2).

7.4 Geometric attacks	
➪ Shift, crop and scale (minimum picture size: 360 x 288)	✔
➪ Bend, shear or rotate by a small amount (up to 2°)	✔ (∇ up to 10°)
7.5 Camera-viewing-display copying technique	✔
7.6 Collusion and collusion-like attacks	✔
8. Watermark editing; for users with secret key	TBD
9. Cascaded watermarking; compatibility of W1 & W2	✔
➪ Number of non-interfering watermarks, invisible in the same picture	At least 2
10. Processing time and delay; real-time embedder & reader	Embedder delay ≤ 80 ms

Bibliography

[1] J. Barda and L. Cheveau: **Eurovision – network security through access control and watermarking**
EBU Technical Review No. 281, Autumn 1999.

[2] J. Barda and L. Cheveau: **Access control and watermarking**
EBU Technical Review No. 282, March 2000.

[3] ITU-R Recommendation BT.500: **Methodology for the subjective assessment of the quality of television pictures**
http://www.itu.int/itudoc/itu-r/rec/bt/500-10.html

Appendix A: Robustness testing of video watermarking systems

(Tests conducted at BBC R&D on behalf of EBU Project Group N/WTM)

The scope of these tests was defined in the EBU Document No. N/WTM 044, NMC 188, *"Watermarking, Call for Systems"*, May 2000. This specified a broadcasting environment consisting of a production level, where watermark W1 was applied, followed by a number of friendly attacks, Ft(X), as shown in *Fig. A1*. The attacks are listed in *Table A1*. This was followed by a contribution-level link (a simulation of the EBU *Eurovision* Network), involving two alternative tests: an MPEG-2 4:2:2 level, Main Profile, 12-frame GOP IBBP encoder at 20.5 and at 8 Mbit/s, followed by a watermark W2. The final stage was a broadcasting-level transmission, consisting of an MPEG-2 ML-MP encoder running at 4 Mbit/s. Tests of detection capability were conducted at the output of each of the above stages, where the material was recorded on DigiBeta machines.

Three final tests involved:

3) ensuring the systems were robust to broadcast coding at 2, 3, 4, 5 and 6 Mbit/s;

4) would pass through a simulation of the BBC's Digital Terrestrial Broadcasting chain, and

5) that the watermark could survive a situation where the broadcast signal is recorded on a single generation of VHS tape, simulating a simple piracy operation.

The contribution-link MPEG codec pair were made by NDS and supplied by the EBU. The broadcast MPEG codec was also simulated by the use of this equipment, running

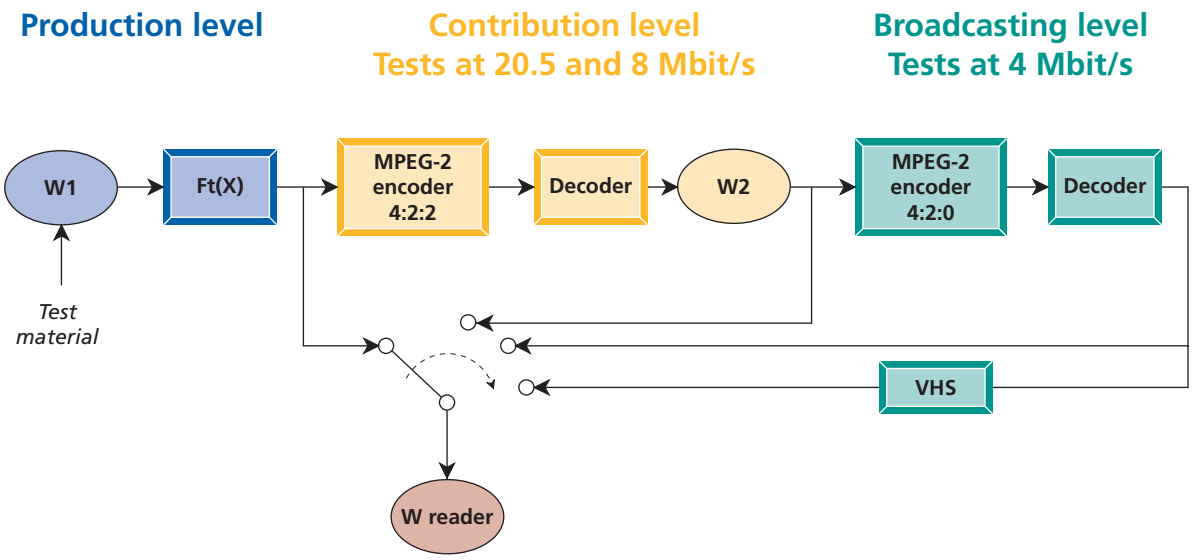


Figure A1
Schematic diagram of the robustness tests (the attacks Ft(X) listed in *Table A1*).

Table A1
Ft(X) – "Production" attacks.

No	Attack	Equipment	Description
1	Digibeta	Sony AVW-A500P	Source format
2	Aspect ratio conversion	DVE: CO6S/514	16:9 to Letterbox on 4:3
3		BBC / Radamec	4:3 to centre of 16:9 with side curtains
4	Shift	DVE	4 pixels right
5			8 pixels left and 4 lines up
6			64 pixels right and 26 lines down
7			200 pixels left and 150 lines down
8			360 pixels right and 288 lines up
9	Scaling	DVE	200% (Zoom in)
10			50% (Zoom out)
11	Bend/shear/rotate un-noticeable amount	DVE	less than 2° rotate
12	Bend/shear/rotate noticeable amount	DVE	less than 10° rotate and shear
13	Re-sampling	D-A + A-D conver-sion	via RGB; also increase in brightness
14	White Noise	TriMedia	Gaussian noise -30 dB, via PAL
15	IMX [a]	c/o Sony Broadcast	MPEG2 4:2:2 50 Mbit/s recorder
16	DV	Sony DV	Low-end production (4:2:0, 25 Mbit/s)
17	PAL and Beta SP	Sony Beta SP	Analogue tape format, PAL input/out-put
18	DVC-Pro or Digital-S	Panasonic DVC Pro	Low-end production (4:1:1, 25 Mbit/s)
19	Beta SX	Sony Beta SX	Low-end production digital tape format
20	Avid	AVR 77	Desktop editor – approx. 3:1 compres-sion
21	Slow Motion	Sony DigiBeta	1/3 rate (0.34)

a. Unfortunately the Sony IMX transfer, although it appeared visually to be fine, proved to have been made on a defective tape machine. A subsequent IMX transfer made by RAI in Italy showed that the original results were suspect, so the results for the IMX attack are incomplete. However, it would be surprising if the results for IMX were significantly different from DigiBeta or Beta SX. Certainly the limited results from the RAI transfer indicate this to be the case.

Tektronix Profile equipment was installed in the broadcast test chain, and so was tested as part of the broadcast system rather than in Ft(X). This operated at a bit-rate of 24 Mbit/s using Motion JPEG compression.

in a conventional ML-MP mode. Transfers and compilation of tapes were done using D1 and DigiBeta recorders. A Tektronix Profile PDR300 video server and a Philips DVS 3112 MPEG-2 encoder were used for the tests that simulated the BBC's digital broadcast chain.

The video test sequence used in these robustness tests was a minute of material, including ten six-second segments of representative material from the RTBF/Talisman test tape and other material, as listed in *Table A2*.

Table A2
Description of the test signals used.

Sequence name and duration		Source	Description
Basketball	6 Sec	MPEG test sequence	Basketball players in stadium with crowd in background. Fast motion. Typical of indoor sports.
Flowers	5 Sec	MPEG test sequence	Highly coloured bed of flowers, with tree moving across foreground and houses in background.
Railings	6 Sec	BBC Widescreen digital production – "Trooping the Colours"	Horseguards riding towards camera, with crowd behind railings in background. Recorded on location on D5 recorder.
Ruck	6 Sec	BBC Widescreen digital production	Close-up of rugby players in a "ruck", surrounded by grass. Recorded on location on D5 recorder.
Belgian Landscape	7 Sec	RTBF/Talisman test tape	Pan across misty landscape ending in sunset
Airport	6 Sec	RTBF/Talisman test tape	Aircraft preparing for, and taking off. Very light background
Satellite Photos	6 Sec	RTBF/Talisman test tape	Three stills
Cabaret	6 Sec	RTBF/Talisman test tape	Singer walks onto stage. Very dark, with large areas of black in the picture
World Cup 94	6 Sec	RTBF/Talisman test tape	Computer-generated Animation
Girl and Rice-field	6 Sec	RTBF/Talisman test tape	Transfer from Film

Whilst the robustness tests were being undertaken, informal subjective opinions were formed of the visibility of the watermarks, and comments on these observations are included.

Purpose of the watermarking robustness tests

A variety of tests need to be undertaken to discover the reliability and robustness of a watermarking system under a variety of "attacks". These attacks can both be in the form of "friendly attacks", such as might be encountered in a normal broadcasting chain, and "hostile attacks" that might be mounted by someone wishing to remove the watermark prior to pirating the video material.

Tests also need to confirm the effect of cascading the watermarks. The addition of a second watermark, normally done using an orthogonal key, should still allow the detection of either watermark independently. This was done by recording watermarked material on DigiBeta, and then playing this material through the watermarking equipment a second time. If the watermark key is the same, but the data payload different, this should render the data payload undetectable. Using different keys, both of the watermarks and their payloads should be recoverable.

Cascading of watermarks from different companies

This was not officially part of the tests, but all parties agreed that whilst the systems were present together it would be sensible to test whether any one system had an effect on the detection of another system's watermark. All combinations of pairs of systems were tested, and in no case did this cause detection problems for any of the systems.