

MPEG-2 4:2:2 and 4:2:0

- comparative subjective tests at low bit-rates

A. Nyberg
SVT

This article reports on subjective test carried out by Swedish Television (SVT) in order to find a possible crossing point where MPEG 4:2:0 becomes superior to MPEG 4:2:2 encoding. The test were performed with the objective of extending the results reported in EBU Technical Review No. 279[1] to bit-rates of 5 Mbit/s and below.

Introduction

The subjective tests described in this report have been performed in order to study the relative performance of MPEG-2 4:2:2P versus MPEG-2 MP@ML encoding profiles. Some sources suggest that a cross-over point exists somewhere in the range 6 to 10 Mbit/s above which 4:2:2 encoding will yield better subjective quality than 4:2:0 encoding.

Previous subjective tests conducted by CBC – with assistance from the EBU and NDS – suggest that 4:2:2 encoding gives better subjective quality than, or the same as, 4:2:0 encoding at bit-rates of 4 Mbit/s and above. The 625-line picture tests carried out at SVT extend these earlier comparisons between the two profiles to cover the range between 2 and 5 Mbit/s. The equipment under test at SVT was the NDS 3000 encoding system, which is used in the *Eurovision* earth stations.

Production of the test material

Test sequences were taken from the EBU Library of Video Sequences. About 20 minutes of the EBU master tape was chosen to be the source material. The chosen sequences included scene 36 (Boy with Toys) up to scene 49 (Family Portrait). Two D5 tape recorders were supplied by the EBU in order to be able to run the test sequences through the MPEG-2 encoding/decoding system without any additional bit-rate reduction due to the recording process.

The test tapes were produced with the test sequences passed through the system at four different bit-rates; 2 Mbit/s, 3 Mbit/s, 4 Mbit/s and 5 Mbit/s. The MPEG-2 encoding employed a GoP length of 12 with an IBBPBBPBBPBB. . . sequence of encoded frames. Sound channels were allocated 128 kbit/s within the available bit-rate.



The digital D5 tape recorders are capable of recording SDI digital component signals at 10-bit resolution. The recording process is bit-transparent, employing no compression in the tape recorders. All connections between the external equipment and the MPEG system under test were at the digital studio standard and 10-bit resolution throughout. Thus, no compression other than the compression employed by the MPEG-2 encoding and decoding was involved in the production of the test material.

Selection of sequences for subjective evaluation

Four well-known scenes were finally selected to be used in the subjective assessment sessions. The four scenes were "Mobile and Calendar", "Renata and Scarf", "Table Tennis" and "Flower Garden". Ten seconds of each scene were chosen to be presented to the observers during the subjective test sessions. The four scenes are regarded as representative samples which are believed to be "critical but not unduly so".

Test method

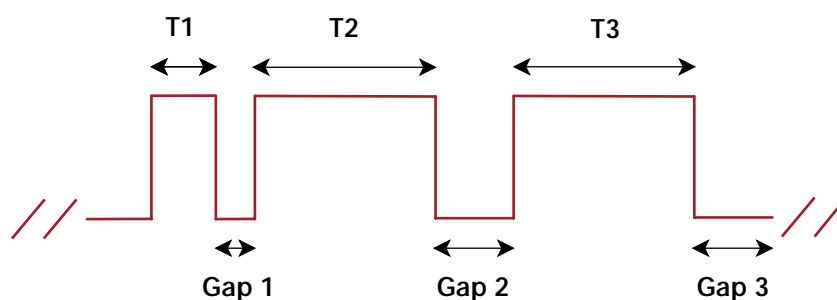
The test method chosen followed ITU-R Recommendation BT.500 [7], paragraph 6.2.4.1 "Adjectival categorical judgement method". The observers were presented with two sequences and were asked to give their judgement of the existence and direction of perceptible differences between the two sequences. The discrete grading scale used (see *Table 1*) is in accordance with ITU-R Recommendation BT.500.

Definition of subjective quality difference	Score
Example A is much worse than B	-3
Example A is worse than B	-2
Example A is slightly worse than B	-1
Example A is the same as B	0
Example A is slightly better than B	+1
Example A is better than B	+2
Example A is much better than B	+3

Table 1
Comparison scale used.

Presentation of the test material

The test session comprised a total of 34 presentations with a structure (cycle) as outlined in *Fig. 1*. Each test condition was presented twice. In the test session, the two first presentations were used as training (stabilizing) sequences and were not part of the evaluation process. The total length of the test session was about 15 minutes.



- T1 = 3s Caption on mid grey (video level -200 mV)
- T2 = 10s Sequence A
- T3 = 10s Sequence B
- Gap1 = 1 s Mid grey
- Gap2 = 2 s Mid grey
- Gap3 = 2 s Voting period, mid grey.

Figure 1
Phases of presentation.



Laboratory set-up

The test sequences were stored on a Panasonic D5 tape recorder and displayed on a Sony PVM 2944 20" monitor. The distance for the assessment was 4H. Editing of the test tape was done using a hard-disk-based system capable of recording the full 10-bit resolution of the test scenes.

Observers and instructions for assessment

Previous to the tests, the assessors were clearly introduced to the method of assessment, the grading scale, the sequence and timing. They were also instructed to give their score according to the meaning of the grading scale. The number of observers was 23 and only integer voting grades were to be used in the voting sheets.

Calculation of results

The mean scores and the 95% confidence limits were calculated for each test condition. Since each test condition was presented twice, the observer's average scores from the two presentations were first calculated and then used as the observer's score for each test condition. The overall test results, with the observer's mean score as a base, was calculated as described in ITU-R Recommendation BT.500, Annex 2.

The test results are given in both *Table 2* and *Fig. 2*. In either case, the results for each of the four individual test sequences are given, followed by an overall mean score, as a function of bit-rate. The calculated mean values with the corresponding 95% confidence limits are depicted in the diagrams. Note that the scale is expanded to show only ± 1 of the total score range of ± 3 .

	Scarf		Table Tennis		Mobile and Calendar		Flower Garden		Overall Mean Value	
	Mean	95% limit	Mean	95% limit	Mean	95% limit	Mean	95% limit	Mean	95% limit
2 Mbit/s	-0.43	0.23	-0.28	0.28	-0.39	0.25	-0.30	0.26	-0.35	0.13
3 Mbit/s	0.04	0.27	-0.57	0.20	0.22	0.23	-0.15	0.26	-0.11	0.12
4 Mbit/s	-0.41	0.22	-0.26	0.22	0.28	0.22	-0.20	0.27	-0.15	0.11
5 Mbit/s	-0.07	0.20	-0.24	0.29	-0.13	0.21	-0.26	0.27	-0.17	0.12

Table 2

Test results at various bit-rates and for different test sequences, showing the mean and overall mean values.

Conclusions

The subjective quality difference between the two systems was judged to be very small. The calculation of mean scores shows only small deviations from zero. For three of the sequences – “Renata and Scarf”, “Table Tennis” and “Flower Garden” – the 4:2:0 encoding did have a tendency to be favoured. For the sequence “Mobile and Calendar”, the 4:2:2 encoding was favoured at bit-rates of 3 Mbit/s and 4 Mbit/s.



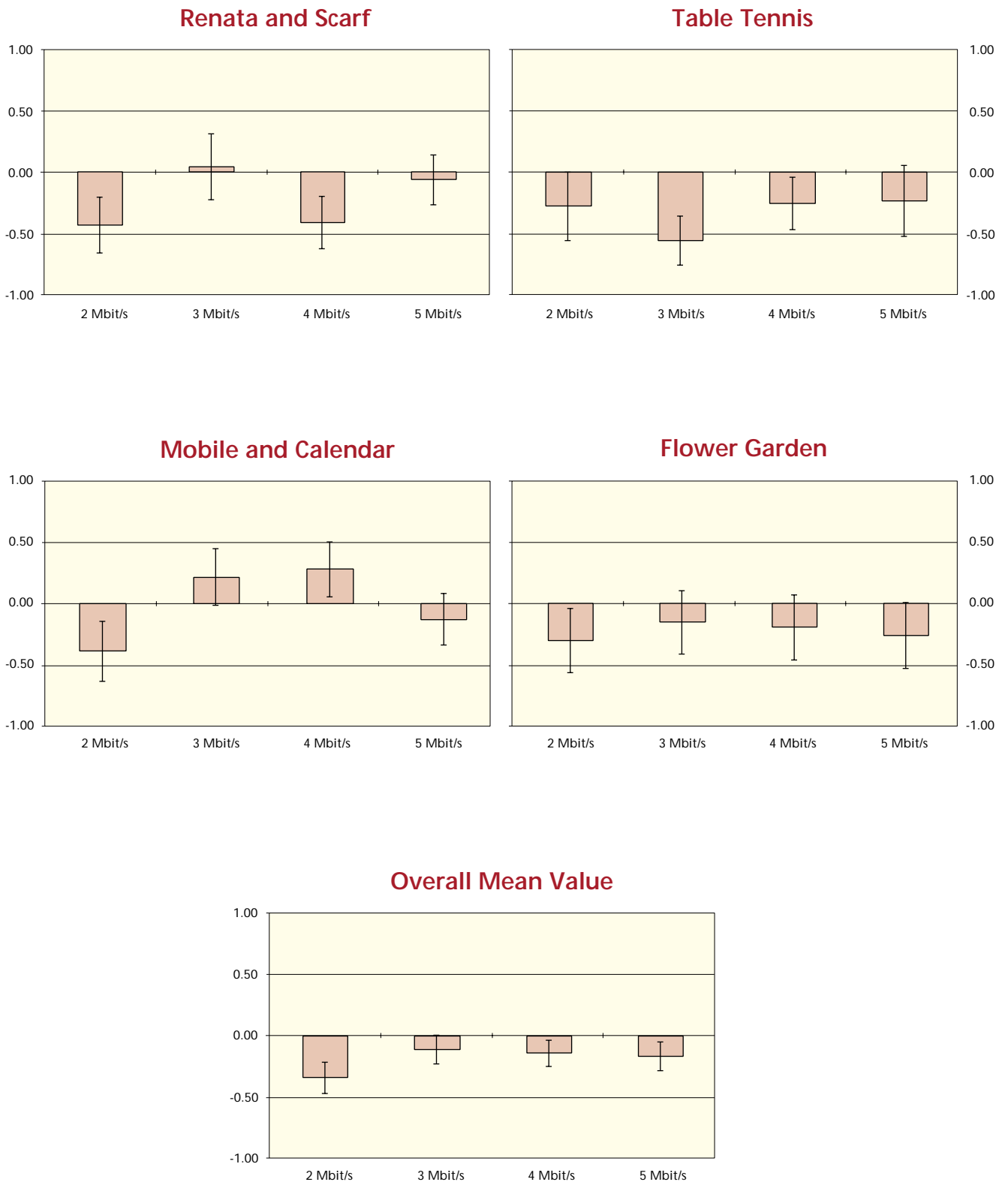


Figure 2
 Test results at various bit-rates and for different test sequences, showing the mean and overall mean values. Note that the vertical scale for each diagram has been expanded to show only ± 1 of the total score range of ± 3 .





Anders Nyberg obtained a master of science degree in electrical engineering at the Royal Institute of Technology, Stockholm, in 1971. After attending the mandatory military services and two years of post-graduate studies on Information Theory, which was combined with teaching duties at the university in Linköping, he joined the development department at Standard Radio and Telephone in Stockholm, working on the design of analogue and digital signal processing circuits and frequency synthesizers for short-wave radio communications equipment.

In 1982 Mr Nyberg joined the research and development department at Sveriges Radio and later, in 1995, the development department of Swedish Television (SVT). During the years 1991-1995, he worked for the Swedish Space Corporation on the Tele-X satellite project, specializing in satellite communications.

Mr Nyberg is currently involved in development activities in transmission technology and communications systems at SVT. He has been involved as a specialist in several EBU subgroups and other activities since 1983 and is currently a member of the EBU Network Management committee, chairman of Project Group N/TBST and vice chairman of Project Group N/NET.

The overall results show that the 4:2:0 encoding system was in general considered to give a small advantage over 4:2:2 at the bit-rates under study and when using the encoding equipment which is used in the *Eurovision* programme contribution network. This is well in line with [8] where a possible cross-over point was expected to be located at about 4 Mbit/s for 625-line systems.

Bibliography

- [7] ITU-R Recommendation BT.500: **Methodology for the subjective assessment of the quality of television pictures**
<http://www.itu.int/publications/itu-r/iturbt.htm>
- [8] L. Cheveau and A. Caruso: **Comparison between 4:2:2P and 4:2:0 – for 525- and 625-line pictures**
 EBU Technical Review No. 279, Spring 1999.

Abbreviations

CBC	Canadian Broadcasting Corporation	ISO	International Organization for Standardization
GoP	Group of pictures	ITU-R	International Telecommunication Union, Radiocommunication Sector
IEC	International Electrotechnical Commission	MPEG	(ISO/IEC) Moving Picture Experts Group
SDI	Serial digital interface	SVT	<i>Sveriges Television</i> (Sweden)