INTERNATIONAL ORGANISATION FOR STANDARDISATION

ORGANISATION INTERNATIONALE DE NORMALISATION

ISO/IEC JTC1/SC29/WG11

CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

ISO/IEC JTC1/SC29/WG11 **N1420**

November 1996

Title: **Overview of the Report on the Formal Subjective Listening Tests of MPEG-2 NBC multichannel audio coding.**

Authors: David Kirby (BBC), Kaoru Watanabe (NHK)

Status: Approved

Abstract

This document presents an overview of the preparations for, and the results of, the subjective listening tests on the MPEG-2 Non-Backwards Compatible (NBC) multichannel audio coding algorithm carried out by the BBC and NHK between 16 September and 11 October 1996. It is derived from the detailed report of the tests [4].

The tests evaluated the following multichannel codecs:

MPEG-2 NBC at 256 kbit/s

MPEG-2 NBC at 320 kbit/s

MPEG-2 NBC low-complexity version at 320 kbit/s

1995 version of MPEG-2 Layer II at 640 kbit/s in a backwards compatible mode.

The test procedure and environments complied with ITU-R Recommendation BS-1116. Listener reliability and test procedure checks were included and a detailed statistical analysis of the results was performed.

The results showed good performance for all of the codecs. The MPEG-2 NBC codec at 320 kbit/s generally performed better than the other codecs and, although not quite transparent for a few test excerpts under these rigorous conditions, it passed the EBU criterion for “indistinguishable quality”. Overall, the performance of MPEG-2 NBC low complexity version at 320 kbit/s was, by a small margin, not quite as good as that of MPEG-2 NBC at 320 kbit/s.

It is worth emphasising that these tests were conducted according to the most rigorous of test methods. Comparisons to other test results using less rigorous methodologies should not be made.

Contents

Abstract 1

1. Introduction 3
   1.1 Background 3
   1.2 Test methodology 3

2. Codecs under test 3
   2.1 Codecs proposed for test 3
   2.2 Status of features used in the MPEG-2 NBC codecs 4

3. Test material 4
   3.1 Selection of test excerpts 4
   3.2 Low anchor presentations 5

4. Experimental design 5
   4.1 Test procedure 5
   4.2 Training 5
   4.3 Listening conditions and test equipment 6

5. Test arrangements at each test centre 6
   5.1 Listening rooms. 6
   5.2 Preparation of test excerpts 6
   5.3 Listening panels 6

6. Statistical analysis and results 6
   6.1 General 6
   6.2 Post-screening to assess listener reliability 7
   6.3 Results for the low-anchor presentations 7
   6.4 Summary of all effects: Analysis of Variance 7
   6.5 Results from the BBC test site 7
   6.6 Results from the NHK test site 9
   6.7 Comparisons of Codecs 10
      6.7.1 MPEG-2 Layer II BC at 640 kbit/s and MPEG-2 NBC at 320 kbit/s 10
      6.7.2 MPEG-2 NBC at 320 kbit/s and MPEG-2 NBC low complexity 11
   6.8 Performance of MPEG-2 NBC at 320 kbit/s according to the EBU definition 11
   6.9 Ranking of the codecs 12

7. Comments on test results. 12
   7.1 Comparison with earlier tests 12
   7.2 Answers to questions in the Test Specification 13
   7.3 Further observations on the tests and the results 13

8. References 14

1. Introduction

1.1 Background

In March 1994, Deutsche Telekom and the BBC reported the results of formal listening tests on the MPEG-2 Backwards Compatible multichannel coding algorithms [1]. Eight codecs were evaluated at that time: six MPEG-2 Backwards Compatible (BC)[1] implementations and two Non-Backwards Compatible (NBC) codecs. The results indicated that none of the codecs tested was acceptable for high quality applications at the tested bitrates. It was also observed that the BC codecs did not perform as well as the NBC codecs at the same bitrate.

As a result of those findings, MPEG decided on two courses of action: firstly, to include, in the proposed MPEG-2 audio standard, additional features which would deliver better audio quality and, secondly, to initiate the development of a Non-Backwards Compatible coding technique. The first of these has led, in stages, to the improved performance of the MPEG-2 BC codecs, reported by earlier subjective tests [2].

The development of the MPEG-2 NBC coding technique has proceeded over the last two years and has reached the stage where formal testing of the multichannel implementation is appropriate. Accordingly, at the July 1996 MPEG meeting, the BBC and NHK were jointly charged to conduct formal subjective tests aimed at quantifying the performance of MPEG-2 Non-Backwards Compatible audio codecs operating in a multichannel mode [3].

During September and October 1996, subjective testing was therefore carried out at the BBC Research and Development Department at Kingswood Warren, UK and at NHK Science and Technical Research Labs, Tokyo, Japan.

This paper provides a summary of the full test report [4] to which the reader should refer for more detailed information and references to the MPEG-2 NBC development work.

1.2 Test methodology

The methodology for these tests was based extensively on the ITU-R Recommendation BS-1116, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems" [5]. This Recommendation had been prepared specifically to highlight any deficiencies of a sound system.

2. Codecs under test

2.1 Codecs proposed for test

The multichannel audio systems under test are all five channel systems with three front channels, left, centre and right, and two surround channels, left surround and right surround. (These tests were carried out without accompanying pictures.)

The following codecs were tested[2]:

- MPEG-2 NBC at 256 kbit/s
- MPEG-2 NBC at 320 kbit/s
- MPEG-2 NBC low-complexity version at 320 kbit/s
- MPEG-2 Layer II at 640 kbit/s in a backwards compatible mode.

The MPEG-2 Layer II BC codec was included to provide a link back to the results of previous subjective tests undertaken in the RACE dTTb project [2] and to provide justification for the existence of the MPEG-2 NBC codec. It was recognised that this MPEG-2 Layer II BC codec implementation might not reflect the current level of performance of the MPEG-2 Layer II BC codec but, as a reasonable approximation, it could be a guide to the relative improvement in performance to be expected from the MPEG-2 NBC coding techniques.

---

[1] In this context, Backwards Compatibility relates to compatibility with MPEG-1 Audio, IS 11172-3

[2] It had been hoped to include Dolby Laboratories AC-3 codec in these tests but this could not be facilitated.

Each of the codecs was submitted as a software implementation.

2.2 Status of features used in the MPEG-2 NBC codecs

The main profile NBC multichannel encoder / decoder is compliant with the MPEG-2 NBC Committee Draft. The main profile has the features: MS stereo, intensity stereo, NEC lossless coding, prediction, temporal noise shaping and dynamic switching of window shape but these features were not all used in each of the embodiments.

All NBC options used the same decoder, which is fully compliant with the MPEG-2 NBC Committee Draft syntax.

The encoders were set to have only the following features active for these tests:

MPEG-2 NBC at 320 kbit/s: this coder used very conservative parameters. Prediction and temporal noise shaping were turned on, but most additional features were switched off to provide the smallest possible change to the RM4 version evaluated earlier in the project [6].

MPEG-2 NBC at 256 kbit/s: this used a combination of prediction, MS stereo coding, temporal noise shaping and dynamic switching of window shape.

MPEG-2 NBC Low Complexity at 320 kbit/s: to reduce the complexity of this implementation, prediction was turned off, however, MS stereo coding, temporal noise shaping (of a lower order than for main profile NBC) and dynamic switching of window shape were activated.

3. Test material

3.1 Selection of test excerpts

A call for suitable five-channel test excerpts with a duration of about 20 seconds was distributed in March 1996 to MPEG members and others working in this field [7]. In total, 94 test excerpts were provided, comprising 66 items made available for earlier tests and 28 new items.
A selection panel was established primarily to identify from these the ten critical excerpts to be used for the tests (their additional tasks are detailed in [4]).

The selection panel consisted of:

- Thomas Buchholz, Deutsche Telekom, TZD-Berlin,
- Kazuho Ono, NHK, Science and Technical Research Labs, Tokyo,
- Andrew McParland and John Fletcher, BBC Research and Development Department.

The panel carried out their work in Listening Room 2 (the same room subsequently used for the formal tests) at the BBC's R&D Department, Kingswood Warren.

The table below lists the ten test excerpts recommended by the panel for these tests.

No	Name	Description
1	pitch_pipe	Pitch Pipe
2	harpsichord	Harpsichord
3	triangle	Triangle
4	cast_pan1	Castanets panned across the front, noise in surround
5	elliot1	Female and male speech in a restaurant, chamber music
6	mancini	Orchestra - strings, cymbals, drums, horns
7	station_master1	Male voice with steam-locomotive effects
8	clarinet_theatre	Clarinet in centre front, theatre foyer ambience, rain on windows in surround
9	thalheim1	Piano front left, sax in front right, female voice in centre
10	glock	Glockenspiel and timpani

Of these items, the panel recommended harpsichord, triangle, Mancini and Thalheim for use as the main items in the training session.

3.2 Low anchor presentations

As the selection panel work proceeded, it became apparent that the tests may not include a sufficient number of mid-quality presentations which are essential in proving the overall sensitivity of the test.

Versions of four of the test items, which were likely to give mid-range quality, were therefore selected from the results of the MPEG ’94 tests [1] as suitable low-anchors. These were chosen on the basis of the mean grades which the items were awarded in those tests: the codec identities and bitrates were unimportant in this choice. The items chosen were: Harpsichord through the MPAC codec (at 320 kbit/s), Mancini through the Layer II codec (at 320 kbit/s) and Pitch pipe and Triangle, both through the AC-3 codec (at 320 kbit/s).

4. Experimental design

The test design followed the ITU-R Recommendation BS-1116 [5] and listeners were asked to judge the all-embracing attribute “Basic Audio Quality” as proposed in that Recommendation.

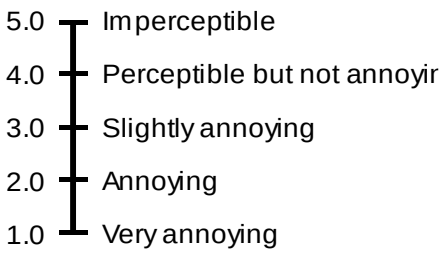
Four weeks were available to carry out the listening tests with at least 20 listeners required at each test site. Generally, three listeners participated every two days. The first half-day was used for training as a group, with the remaining one and half days available for the grading phase.

4.1 Test procedure

The tests used the “triple stimulus/hidden reference/double blind” method and were carried out by the listeners individually over several sessions, each of which took about 25 minutes. The listener could switch freely between the presentations “Reference”, “A” and “B”, where “A” and “B” were the processed version and the hidden reference, randomly allocated from one trial to the next. (The allocation was known neither to the listener nor to the test supervisor, hence the term “double blind”.) Each test excerpt could be repeated as often as the listener wished. The listener was asked to judge the “Basic Audio Quality” of the “A” and “B” versions in each trial and any difference from the reference was to be considered as an impairment. The order of the test presentations and the position of the hidden reference were randomised for each listener.

Each listener graded the perceived differences using the ITU-R five-point impairment scale shown alongside.

The listeners were asked to input their grades to an accuracy of one decimal place. At least one grade of “5.0” had to be given for each trial, since one of “A” or “B” was the hidden reference.



4.2 Training

The morning of the first day was used for a joint training session involving the three listeners for that two-day session. This allowed them to become familiar with the test procedure, assist each other in identifying coding artefacts and to become more experienced listeners. They were guided during this training by a test supervisor. However, to maintain the blind nature of the tests, both the test supervisor and the listeners were unaware of the identities of the codecs and bitrates being used.

In the early stages, the training made use of bitrates lower than those used in the tests to make the impairments clearer for listeners, particularly for those less familiar with this type of artefact.

Throughout the training and the tests, the listeners were asked not to discuss the grades they would award as this was required to be an individual subjective judgement for each of them.

The steps taken in the training phase were similar at both test centres. Firstly, an initial impression of the ten test excerpts and coding artefacts was demonstrated by replaying the reference version and a coded version of each. Then, each of the four training excerpts (a subset of the ten test excerpts), coded with one of the codecs at a lower bitrate, was replayed in reference and coded forms and extensively assessed and discussed. Following this, the same four training excerpts at

both bitrates and through each codec were auditioned and discussed. Finally, the remaining six excerpts were again demonstrated in coded and reference forms.

After this training, the listeners each carried out a 'mini-test' to allow practice with the test control system and also to accustomise themselves to listening individually. This used the four main training items arranged as a randomised Ref/A/B test. Each listener was allowed about ten minutes to do this test and was advised that the results were unimportant and would not be used.

4.3 Listening conditions and test equipment

ITU-R Recommendation BS-1116 [5] defines specific requirements for the listening conditions to ensure comparable and reliable results of subjective assessments of sound systems. These cover:

- the acoustical characteristics of the listening room and the sound field therein,
- the arrangement of the monitoring loudspeakers in the listening room,
- the location of the listening positions for the test.

The listening arrangement specified in this Recommendation was used at both test sites, with the listener in the "Centre" or "Reference Listening Position".

5. Test arrangements at each test centre

5.1 Listening rooms.

At the BBC, Listening Room 2 at Research and Development Department, Kingswood Warren, was used for these tests. This room was used for the previous multichannel tests: the MPEG tests in 1994 [1] and the RACE dTTb tests in 1996 [2]. Although slightly smaller than recommended in the ITU-R Recommendation BS-1116 [5] for multichannel tests, in most other aspects the requirements are met.

At NHK, Listening Room B268 at the Science and Technical Research Laboratories, Tokyo Japan, was used. The requirements of BS-1116 are met in most respects.

5.2 Preparation of test excerpts

At the BBC, the ten test excerpts for the four codecs, plus the four low anchor excerpts, were compiled into eight blocks, each containing either five or six excerpts, with codecs and excerpts in a randomised order. Each excerpt was recorded, sample aligned with its corresponding reference, with nine repetitions, on to two Tascam DA88 tapes.

The test excerpts were received by NHK on Exabyte tape from the BBC. These were loaded on to a SonicStudio audio editor from which they could be replayed, synchronised with the reference recording, in any order.

5.3 Listening panels

All the listeners were professionally involved in audio work: 32 participated at the BBC and 24 at NHK. The majority had a background in sound production, whilst the remaining listeners were involved in audio engineering. Potential listeners affiliated to the codec proponents were excluded.

6. Statistical analysis and results

6.1 General

Throughout the statistical analysis[3] 'diffgrades' are used; these are calculated, from each trial, as the grade awarded to the coded version minus the grade awarded to the reference. Thus, for example, an impairment grade of 4.0 ("Perceptible but not annoying") awarded *to the coded version* becomes a diffgrade of -1.0, whilst a grade of 5.0 ("Imperceptible") gives a diffgrade of 0.0. This relationship must be borne in mind when relating the numerical results to the level of artefacts perceived.

[3] These analyses were performed by Dr Alan Kimber and Mr Peter Williams of the University of Surrey, UK.

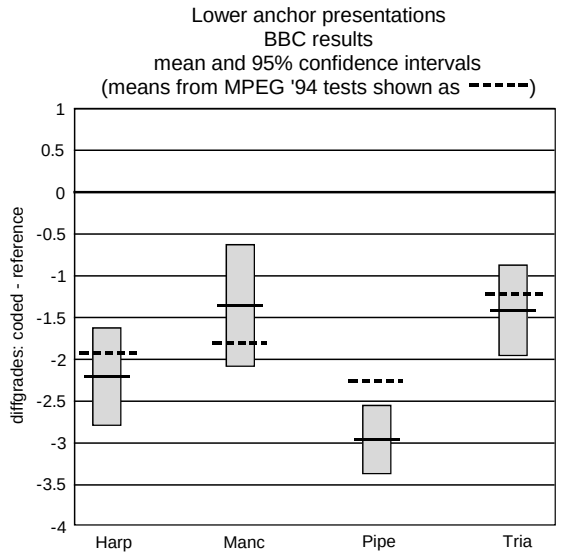
6.2 Post-screening to assess listener reliability

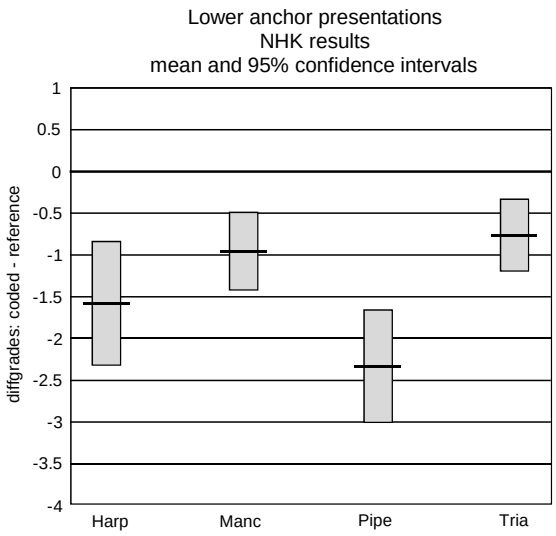
As suggested in the ITU-R Recommendation BS-1116 [5], a post-screening of all the listeners was carried out by using a one-sided t-test at a confidence level of 95%. The probability of accepting a subject who, on average, was unable to detect the coded version, is then 0.05 at most. In addition to this procedure, a Wilcoxon test was also applied to assess reliability.

This analysis showed that 23 out of the 32 BBC listeners and 16 out of the 24 NHK listeners were judged to be reliable in these tests. The rest of the analysis used data from only these listeners.

6.3 Results for the low-anchor presentations

As the additional four low-anchor presentations originated from different codecs, and gave data for only 4 of the ten test excerpts, their data was removed from the remainder of the analysis. However, the results for these presentations were calculated separately and are shown below. For the BBC results, the corresponding mean diffgrades from the MPEG '94 tests [1] are also included.





From the above BBC results for the low-anchor presentations, it can be seen that the means from the previous tests for three of the four excerpts lie within the 95% confidence intervals of these tests.

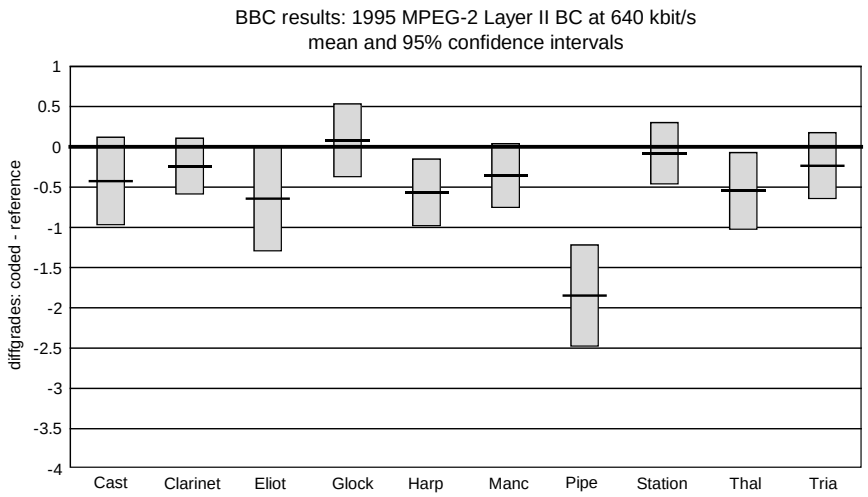
These results show that the test arrangements, *i.e. the listening conditions and listeners together*, at both test sites, were able to reveal artefacts in a range below transparency: thus the validity of the test arrangements is confirmed.

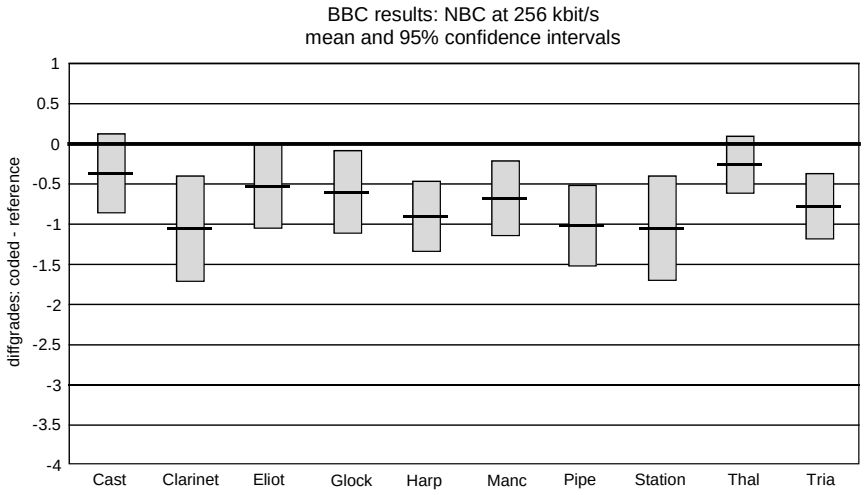
6.4 Summary of all effects: Analysis of Variance

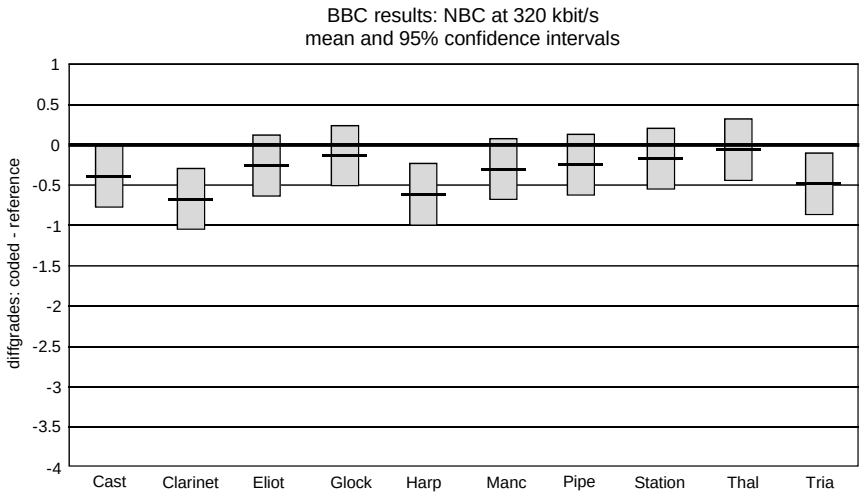
Using the data from the reliable listeners, a three-way ANOVA was performed with main effects of "Site", "Codec" and "Item". This showed that the Site effect was significant and so the data from both sites could not be combined. The remainder of the analysis was, therefore, performed separately for the BBC and NHK data and two sets of results are presented.

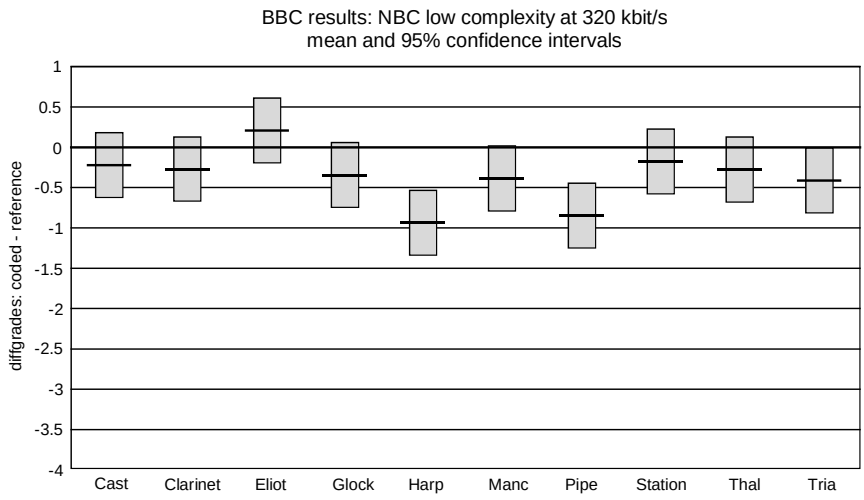
6.5 Results from the BBC test site

The following four diagrams show the estimated means and two-sided 95% confidence intervals for each codec and excerpt at the BBC site. For the first two codecs, the data fails to meet certain statistical assumptions and so the confidence intervals are calculated individually for each excerpt. For the latter two diagrams the assumptions are valid and so the confidence intervals are calculated from the ANOVA and hence, are of equal length within each group.


BBC results: 1995 MPEG-2 Layer II BC at 640 kbit/s
mean and 95% confidence intervals
diffgrades: coded - reference
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
Cast
Clarinet
Eliot
Glock
Harp
Manc
Pipe
Station
Thal
Tria

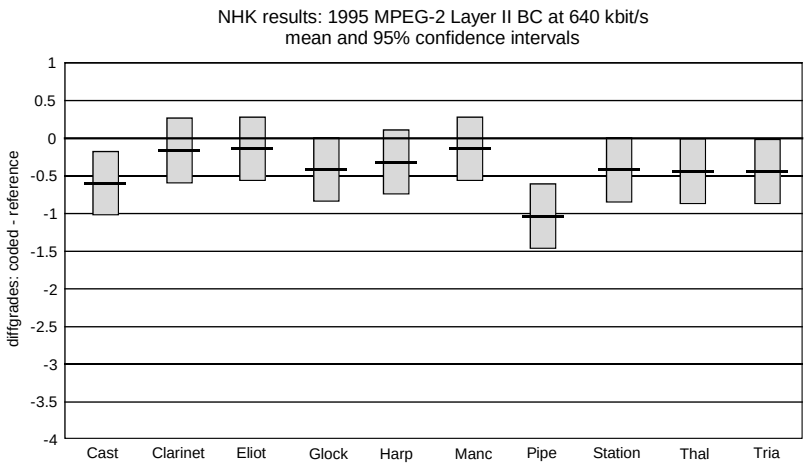

BBC results: NBC at 256 kbit/s
mean and 95% confidence intervals
diffgrades: coded - reference
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
Cast
Clarinet
Eliot
Glock
Harp
Manc
Pipe
Station
Thal
Tria

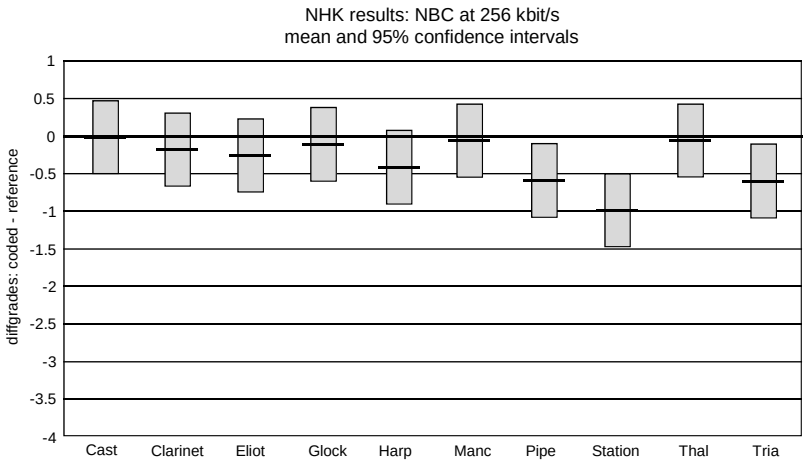

BBC results: NBC at 320 kbit/s
mean and 95% confidence intervals
diffgrades: coded - reference
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
Cast
Clarinet
Eliot
Glock
Harp
Manc
Pipe
Station
Thal
Tria

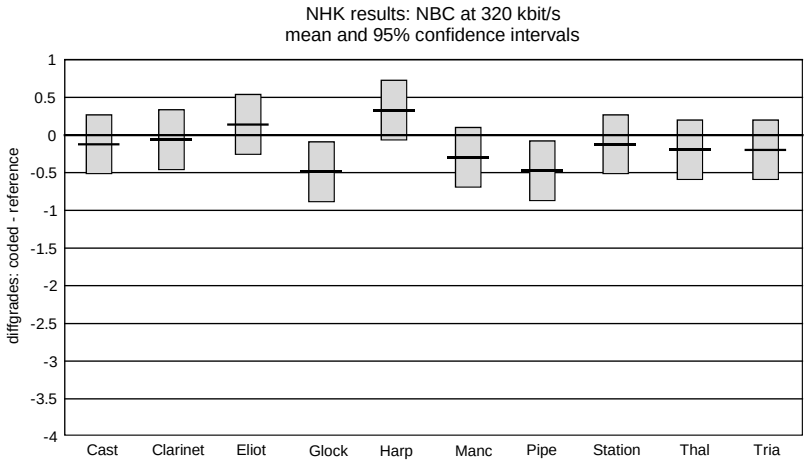


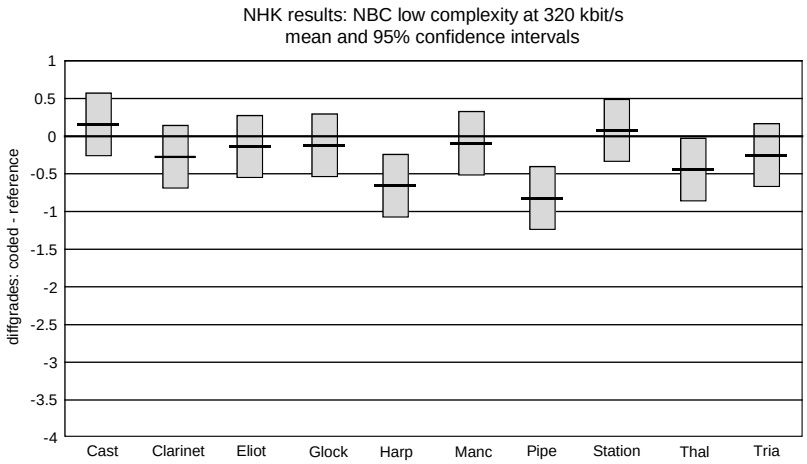
6.6 Results from the NHK test site

The following four diagrams show the estimated means and two-sided 95% confidence intervals for each codec and excerpt. In each case, the confidence intervals are calculated from an ANOVA and, hence, are of equal length within each group.







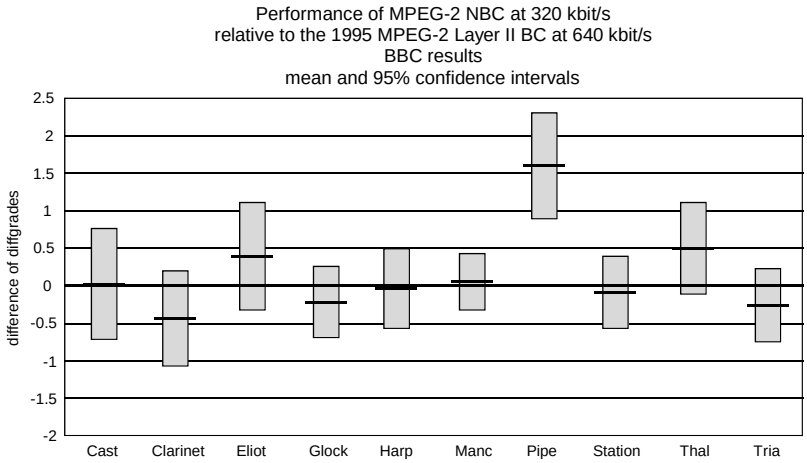


6.7 Comparisons of Codecs

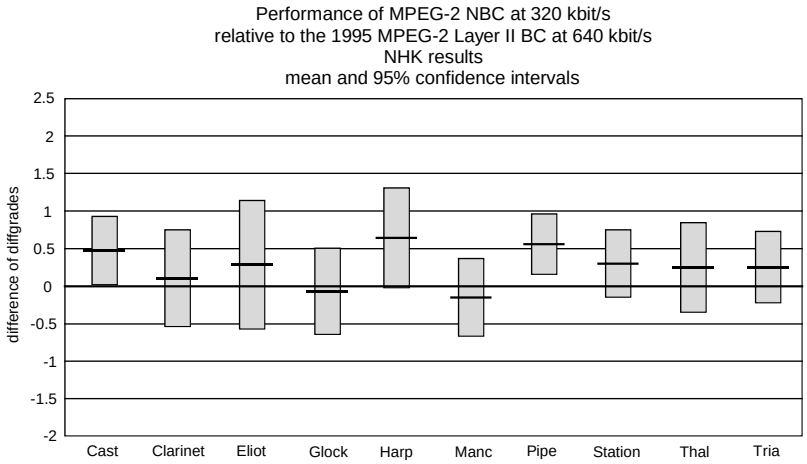
6.7.1 MPEG-2 Layer II BC at 640 kbit/s and MPEG-2 NBC at 320 kbit/s

In order to determine if significant differences exist between the MPEG-2 Layer II BC codec at 640 kbit/s[4] and the MPEG-2 NBC codec at 320 kbit/s, two ANOVAs (one for each site) were performed on the data for these two codecs. These both revealed differences between these two codecs. This can be seen from the diagrams in Sections 6.5 and 6.6, with the "NBC at 320 kbit/s" diffgrades generally out-performing (i.e. being closer to zero than) the "1995 Layer II at 640 kbit/s" diffgrades.

The differences between the diffgrades for these two codecs were calculated, item by item, and these are shown below for each test site (a positive difference value indicating that the MPEG-2 NBC codec was awarded a better diffgrade than the MPEG-2 Layer II BC codec).

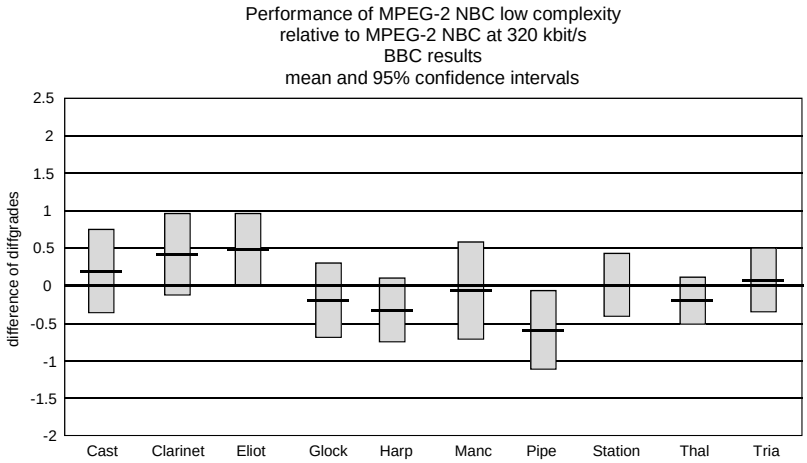


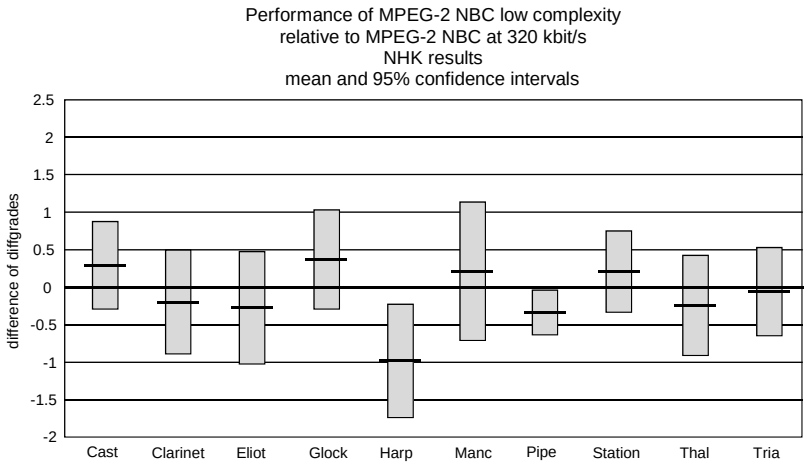
[4] Note, the reader is reminded that these test results relate to the 1995 version of MPEG-2 Layer II BC and do not reflect any subsequent enhancements that may have occurred.



6.7.2 MPEG-2 NBC at 320 kbit/s and MPEG-2 NBC low complexity

As the performance of the MPEG-2 NBC low complexity implementation appears to be significantly better than originally suggested, a comparison with MPEG-2 NBC at 320 kbit/s has been carried out. The differences between the diffgrades for these two codecs are shown below (a positive difference value indicating that the MPEG-2 NBC low complexity codec was awarded a better diffgrade than the MPEG-2 NBC codec at 320 kbit/s).





6.8 Performance of MPEG-2 NBC at 320 kbit/s according to the EBU definition

The test specification [3] asks if the performance of the NBC codec at 320 kbit/s achieved ‘indistinguishable quality’ according to the EBU definition [8].

A detailed analysis showed that only the MPEG-2 NBC codec at 320 kbit/s passes the EBU criterion at both sites. The MPEG-2 NBC low complexity codec passes at NHK but is borderline at

the BBC. The other codecs, MPEG-2 Layer II BC at 640 kbit/s and MPEG-2 NBC at 256 kbit/s, fail at both sites.

It should be pointed out that at neither site were there 40 or more subjects (as laid down by EBU).

6.9 Ranking of the codecs

To determine if a relative ranking of the codecs could be established, two analyses were carried out. A Least Significant Difference analysis of codec means by site did not give a particularly clear picture. A simple method of comparison was then performed based on the number of items for each codec at each site for which the 95% confidence interval for mean diffgrade (a) contained 0 and (b) contained -1 or less This gave the following results:

Codec /Site	Number with 0 in Conf. Int.	Number with -1 or less in Conf. Int.
Layer II at 640 / BBC	6	3
NBC at 256 / BBC	2	8
NBC at 320 / BBC	6	1
NBC lc at 320 / BBC	7	2
Layer II at 640 / NHK	5	2
NBC at 256 / NHK	7	3
NBC at 320 / NHK	8	0
NBC lc at 320 /NHK	7	2

This, perhaps, indicates a rough ordering of codecs: "NBC at 320" and "NBC low complexity at 320" (best), followed by codec "Layer II at 640" and finally codec "NBC at 256" for the BBC site.

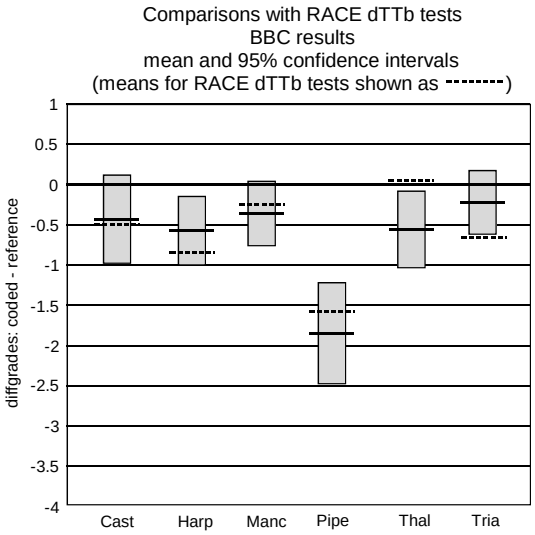
For the NHK site this ordering is: codec "NBC at 320" (best), followed by codec "NBC low complexity at 320", then codecs "Layer II at 640" and "NBC at 256".

Interestingly, these results agree quite well with the EBU criteria presented in Section 6.8.

7. Comments on test results.

7.1 Comparison with earlier tests

The BBC results for the low anchor presentations (Section 6.3) show consistency between these and the MPEG ‘94 tests [1]. For a direct comparison with the RACE dTTb tests [2], the results for the 1995 MPEG-2 Layer II BC codec at 640 kbit/s are shown in the diagram below for both tests.



These show good agreement with four of the six mean values from the dTTb tests lying within the confidence intervals for these tests. The apparent differences between the means for the Thalheim and Triangle items were further investigated by applying a two-sided t-test. This indicated that, at the 95% confidence level, there is no significant difference between the means from each test.

7.2 Answers to questions in the Test Specification

The test specification [3] poses various questions to be investigated by the statistical analysis. These are answered as follows:

1) Are the listeners' results reliable, i.e. distinguishable from random votes?
   An assessment of listener reliability has been performed and only reliable data has been used in the analysis (see Section 6.2).
2) Does the test methodology allow meaningful conclusions to be drawn from these results?
   Yes. Furthermore, performance below the level of transparency could be detected and these tests also appear to be reasonably consistent with earlier tests.
3) Is there any distinction between the two test sites?
   Yes; a three-way ANOVA revealed differences between the sites (see Section 6.4).
4) Is the performance of MPEG-2 NBC at the default bitrate *[320 kbit/s]* equal to or better than the performance *[of the 1995 version]* of MPEG-2 BC Layer II at 640 kbit/s?
   Differences between these codecs were revealed. Generally, the performance of the MPEG-2 NBC codec at 320 kbit/s appears to be better.
5) How does the performance of the codecs vary with programme items?
   Sections 6.5 and 6.6 of this report show the performance of each codec for each of the programme items.
6) Is the performance of the coding of NBC at the default bitrate *[320 kbit/s]* distinguishable from the original signal?
   The diagrams shown in Sections 6.5 and 6.6 indicate that the MPEG-2 NBC codec at 320 kbit/s is distinguishable from the original signal for some excerpts.
7) Is the performance of NBC at the default bitrate *[320 kbit/s]* achieving 'indistinguishable quality' in the EBU definition [8] of that phrase?
   Yes, at both test sites. (However, in each case, fewer than the recommended 40 listeners participated). See Section 6.8.
8) What is the relative ranking of the codecs tested?
   A clear ranking of the codecs is difficult to determine as their grouping differs between the test sites. However, generally, MPEG-2 NBC at 320 kbit/s and MPEG-2 NBC low complexity performed better than the 1995 version of MPEG-2 Layer II BC at 640 kbit/s and MPEG-2 NBC at 256 kbit/s; see Section 6.9 for further details.
9) Are there any other features from the data that should be reported?
   A comparison of MPEG-2 NBC at 320 kbit/s and MPEG-2 NBC low complexity has been included, see Section 6.7.2.

7.3 Further observations on the tests and the results

- The results for all the codecs show very good performance. During the tests, most subjects found it necessary to listen to each trial many times because of the difficulty in identifying the coded version.
- All the variants of MPEG-2 NBC coding which were tested, achieved approximately the same 5-channel performance or better at half the bitrate of the 1995 version of the MPEG-2 Layer II BC codec.
- At both test sites, the MPEG-2 NBC codec at 320 kbit/s achieved diffgrades better than -0.7 for all of the test excerpts (i.e. better than grade 4.3 on the impairment scale). Only two of these excerpts, clarinet and harpsichord at the BBC site, gave mean diffgrades worse than -0.5 and only one excerpt, clarinet, gave rise to a 95% confidence interval which crossed below the diffgrade value of -1.0.
- The MPEG-2 NBC low complexity implementation at 320 kbit/s achieved mean diffgrades at both test sites better than -1.0, i.e. better than grade 4.0 on the impairment scale.
- With the implementations tested, MPEG-2 NBC low complexity at 320 kbit/s is only marginally worse than MPEG-2 NBC at 320 kbit/s.

- Care should be exercised when comparing the performance of the different implementations of the MPEG-2 NBC codecs as they had different features enabled in addition to the differing bitrate or level of complexity.
- Where test stimuli, low anchors or MPEG-2 Layer II BC, had been previously assessed in earlier tests, the results from this series of tests are very similar to those previously published.
- Further evaluations of the MPEG-2 NBC coders may be warranted once further coding optimisation has been carried out.
- No assessments have yet been reported on the two-channel stereo performance of the MPEG-2 NBC codecs. If two-channel reproduction is to be achieved by simulcasting using an existing stereo coder, then results from stereo coding tests can be assumed to be relevant, but the bitrate will increase accordingly. If the stereo version is to be created by downmixing of the 5-channels delivered by MPEG-2 NBC coding, then the bit rate will be as reported here, but subjective assessments of the stereo performance should be made.

8. References

1. Feige, F. & Kirby, D.G., 1994, “Report on the MPEG/Audio multichannel formal subjective listening tests”. MPEG document ISO/IEC JTC1/SC29/WG11/N0685. March 1994.
2. Feige, F. & Kirby, D., 1996. “MPEG-2 Backwards compatible codecs Layer II and Layer III: RACE dTTb listening test report”. ISO/IEC JTC1/SC29/WG11/N1229. March 1996.
3. Feige, F., Meares, D. J., “Specification of MPEG-2 Audio NBC Formal Tests”. MPEG document ISO/IEC JTC1/SC29/WG11/N1281. July 1996.
4. Kirby, D. G. & Watanabe, K., 1996. “Report on the formal subjective listening tests of MPEG-2 NBC multichannel audio coding”. ISO/IEC JTC1/SC29/WG11/N1419. November 1996
5. ITU-R Recommendation BS.1116 “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems”. Geneva (1994)
6. Lueck, C., et al, 1996. “NBC Reference Model 4 stereophonic and multichannel subjective tests: overall results”. ISO/IEC JTC1/SC29/WG11/N1280. July 1996
7. Fielder, L., 1996. “Call for multichannel audio test sequences”. ISO/IEC JTC1/SC29/WG11/N1203. March 1996.
8. EBU, 1991. “Basic audio quality requirements for digital audio bit-rate reduction systems for broadcast emission and primary distribution”. CCIR document number TG 10-2/3, 28 October 1991