INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11/N5571

March 2003, Pattaya, Thailand

Source: **Audio**
Status: **Approved at the 64th Meeting**
Title: Report on Informal MPEG-4 Extension 1 (Bandwidth Extension) Verification Tests

1 Introduction

In mid-1999 International Standard ISO/IEC 14496-3, MPEG-4 Audio Version 1 issued and in early 2000 ISO/IEC 14496-3 / AMD1, MPEG-4 Audio Version 2 issued. Since then, numerous tests have been conducted by MPEG to verify that the MPEG-4 standard contains state of the art technology. However, WG11 is always interested in new developments which may provide improvements over the existing MPEG-4 standard and which may lead to extensions of MPEG-4 or to new work items.

For this reason, at the 53rd MPEG meeting, in Beijing, MPEG issued a Call for Evidence Justifying the Testing of Audio Coding Technology (N3641). Evidence submitted in response to the Call was examined at the 55th MPEG meeting, in Pisa, and it was determined that there was technology that might improve upon the MPEG-4 standard.

Based on the results of the Call for Evidence, work was begun in WG11 to standardize technology for an MPEG-4 Bandwidth Extension tool that could be applied to general audio signals.

2 Acceptance criteria

The performance of the new technology would be measured via a subjective test, carried out prior to the Final Draft International Standard stage of the standardization process. The acceptance criteria (enumerated below) must be met in order for the technology to progress to International Standard.

Using the MUSHRA test methodology, and with the target bit-rate for the Bandwidth Extension coder set to 24 kbit/s per channel for general audio signals, the Bandwidth Extension technology shall satisfy the following two criteria.

1. With the Bandwidth Extension operating at the target bit-rate and MPEG-4 operating at 25% higher bit-rate, the Bandwidth Extension technology shall have a mean score that is comparable to or better than the mean score of MPEG-4.
2. With both coders operating at the target bit-rate, none of the items shall be worse in a statistical sense for the Bandwidth Extension technology.

As a result of optimizing, the quality of the backward compatible part might be less than that of a regular encoder. The quality of the core coder shall be compared to MPEG-4 operating at a bit-rate 25% lower than the target bit-rate.

3 Codecs under test

Two versions of Bandwidth Extension technology have been tested:

- High Quality SBR
- Low Power SBR

They are tested using both Mono and Stereo signals, as detailed in the following table:

Table 1 - Modules included in Mono and Stereo Tests

Mono Signals	**Label**	**Bitstreams**	**Decoder and *.wav files**
Hidden reference	H-Ref-Org		
Hidden reference bandlimited to 7 kHz	H-Ref-7		
Hidden reference bandlimited to 3 kHz	H-Ref-3		
HQ SBR at 24 kb/s	HQ-SBR-24	Coding Technologies	Coding Technologies
LP SBR at 24 kb/s	LP-SBR-24	bitstr. identical to above	Coding Technologies
MP4 AAC LC at 30 kb/s	MP4-AAC-30	Same as CfP	Same as CfP
MP4 AAC LC at 24 kb/s	MP4-AAC-24	Same as CfP	Same as CfP
AAC only part of HQ (or LP) SBR at 24 kb/s	BC-AAC-24	bitstr. identical to HQ-SBR-24	Coding Technologies

Stereo Signals	**Label**	**Bitstreams**	**Decoder and *.wav files**
Hidden reference	H-Ref-Org		
Hidden reference bandlimited to 7 kHz	H-Ref-7		
Hidden reference bandlimited to 3 kHz	H-Ref-3		
HQ SBR at 48 kb/s	HQ-SBR-48	Coding Technologies	Coding Technologies
LP SBR at 48 kb/s	LP-SBR-48	bitstr. identical to above	Coding Technologies
HQ SBR at 32 kb/s	HQ-SBR-32	Coding Technologies	Coding Technologies
LP SBR at 32 kb/s	LP-SBR-32	bitstr. identical to above	Coding Technologies
MP4 AAC LC at 60 kb/s	MP4-AAC-60	Same as CfP	Same as CfP
MP4 AAC LCat 48 kb/s	MP4-AAC-48	Same as CfP	Same as CfP
AAC only part of HQ (or LP) SBR at 48 kb/s	BC-AAC-48	bitstr. identical to HQ-SBR-48	Coding Technologies
AAC only part of HQ (or LP) SBR at 32 kb/s	BC-AAC-60	bitstr. Identical to HQ-SBR-32	Coding Technologies

The MP4-AAC reference bitstreams were identical to those used for the "Call for proposals" test. The SBR enhanced bitstreams and the decoded wav-file outputs, as well as the SBR enhanced SBR decoders, were provided by Coding Technologies. No MP4-AAC reference streams at 25% lower bitrate than the bitrate of the SBR enhanced bitstreams were available. Hence, the final comparison could not be made.

The quality difference between the High Quality SBR and the Low Power SBR has been evaluated in a core experiment. The differences were established in a CMOS test, were the High Quality SBR performed better than the Low Power SBR.

4 Test material

The following test material was used:

- speech signals [es*]
- single instruments (monophonic, i.e. one note sounding at a time) [si*]
- simple sound mixtures (material with. several notes sounding at a time) [sm*]
- complex sound mixtures [sc*]

Test Item	Description
es01	vocal (Suzanne Vega)
es02	German speech
es03	English speech
si01	Harpsichord
si02	Castanets
si03	pitch pipe
sm01	Bagpipes
sm02	Glockenspiel
sm03	Plucked strings
sc01	Trumpet solo and orchestra
sc02	Orchestral piece
sc03	Contemporary pop music

This material is available at 48 kHz sampling rates. Bitstream providers can create other sampling rates by using the **ResampAudio** sample rate conversion tool:
http://www.tnt.uni-hannover.de/soft/audio/packages/afsp/

5 Listening test methodology

The “Subjective assessment of sound quality” (MUSHRA) test methodology was used. (This method is a standard ITU-R Recommendation BS.1534, available at http://ecs.itu.ch)

The following stimuli will be used as references throughout the tests:

1. full bandwidth reference
2. full bandwidth hidden reference
3. low pass filtered hidden reference (7 kHz)
4. low pass filtered hidden reference (3.5 kHz)

The listening test design is defined as follows:

- at least a total of 8 listeners
- expert listeners are preferred
- training with the corresponding selected training items
- computer-based MUSHRA presentation will be used

- headphones STAX (preferred STAX LAMBDA PRO)
- one listener at a time due to open headphones
- the test shall be divided in sections of approx. 20 minutes length
- the test will be preceded by a training

5.1 Training of subjects

The first step of training is to listen to the training items in order to become familiar with the nature of the artifacts. The subjects can discuss the perceived artifacts, but subjects are not allowed to talk about specific grades in order to avoid bias in individual grading. The randomization of the order of presentation of the training items and the number of repetitions of the items was at the discretion of each listening test site.

The second step of the training is to run a dummy grading of the training items using the grading facility (paper sheet or on-screen display) to become familiar with this tool for the subsequent grading phase.

6 Test Sites

The following sites performed the listening tests:

Site	**Listeners**
Coding Technologies	9
Matsushita	4
NEC	4
PSL	3
Total	20

7 Tests Results

The results of the listening tests are presented in two sections, one for mono signals, and the other for stereo signals. The acceptance criterion are assessed via two methods:

- A plot of grand mean and 95 % confidence interval of the subjective grade of each coder under test, as averaged over all test sites, all listeners and all test excerpts.
- A table of mean score and 95 % confidence interval of the subjective grade of each coder under test for each test excerpt, as averaged over all test sites and all listeners.

In the same zip archive as this document are two Excel spreadsheets, “BWE Mono Mushra” and “BWE Stereo Mushra” that contain all data, pivot tables, plots and tables used in the analysis of the listening tests.

7.1 Monophonic Signals

The plot in Figure 1 shows the results of the subjective test of monophonic signals. The horizontal tick is the mean value, the vertical tick is the 95% confidence interval. It shows that the Bandwidth Extension technology operating at 24 kb/s, indicated by labels HQ-SBR-24 and LP-SBR-24, each have a grand mean score that is comparable to or better than the grand mean score of MPEG-4 AAC operating at 30 kb/s, indicated by label MP4-AAC-30. Hence the first acceptance criterion for both systems is passed.

Table A.1 lists the lower and upper bounds of the 95% confidence interval (CI) of the mean score for each test excerpt. Careful examination of the table indicates that, with all coders operating at the target bit-rate of 24 kb/s, none of the items are worse in a statistical sense for the developed

technology as compared to MPEG-4 AAC (in other words, the 95% CI upper limit of developed technology is never less than the lower limit of MPEG-4 AAC).

Hence, both the High-Quality Bandwidth Enhancement technology (HQ-SBR-24) and the Low Power Bandwidth Enhancement technology (LP-SBR-24) pass the acceptance criterion.

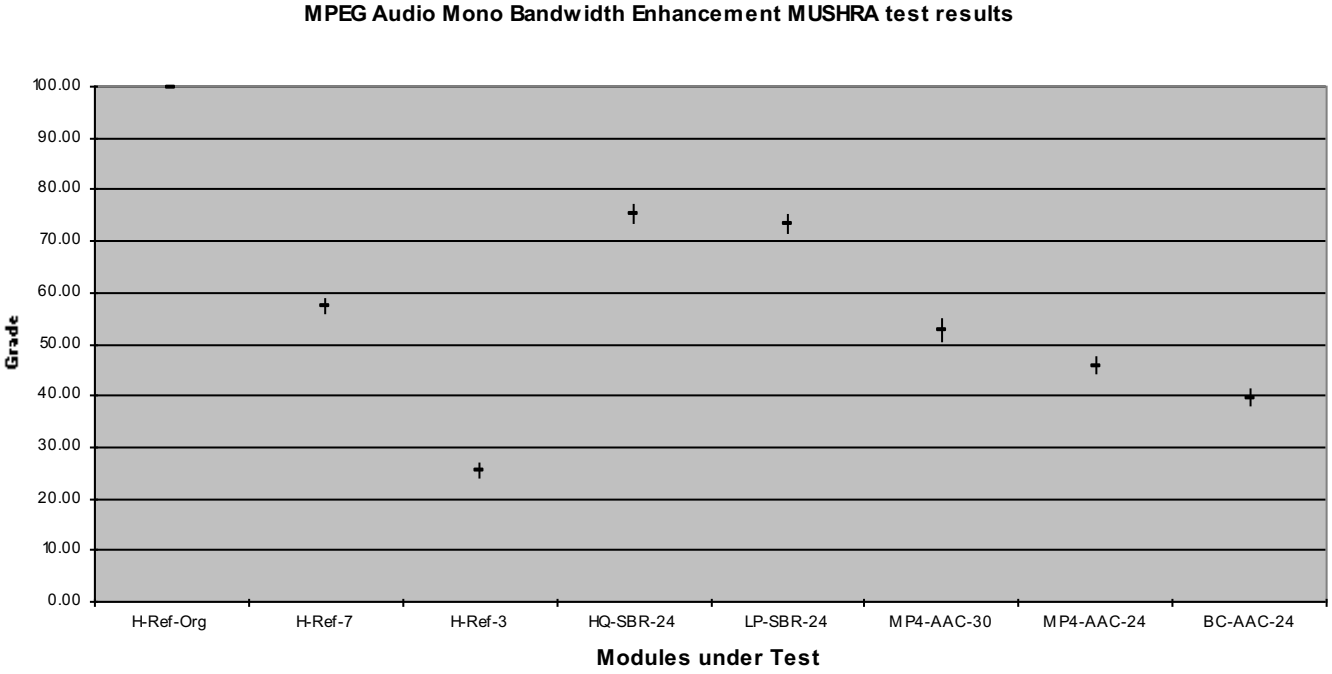

Figure 1- Mono Test Results. See Table 1 "Label" column for key to horizontal axis labels.

7.2 Stereophonic Signals

The plot in Figure 2 shows the results of the subjective test of stereophonic signals. The horizontal tick is the mean value, the vertical tick is the 95% confidence interval. It shows that the Bandwidth Extension technology operating at 48 kb/s, indicated by labels HQ-SBR-48 and LP-SBR-48, each have a grand mean score that is comparable to or better than the grand mean score of MPEG-4 AAC operating at 60 kb/s, indicated by label MP4-AAC-60. Hence the first acceptance criterion for both systems is passed.

Table A.2 lists the lower and upper bounds of the 95% confidence interval (CI) of the mean score for each test excerpt. Careful examination of the table indicates that, with all coders operating at the target bit-rate of 48 kb/s, none of the items are worse in a statistical sense for the developed technology as compared to MPEG-4 AAC (in other words, the 95% CI upper limit of developed technology is never less than the lower limit of MPEG-4 AAC).

Hence, both the High-Quality Bandwidth Enhancement technology (HQ-SBR-48) and the Low Power Bandwidth Enhancement technology (LP-SBR-48) pass the acceptance criterion.

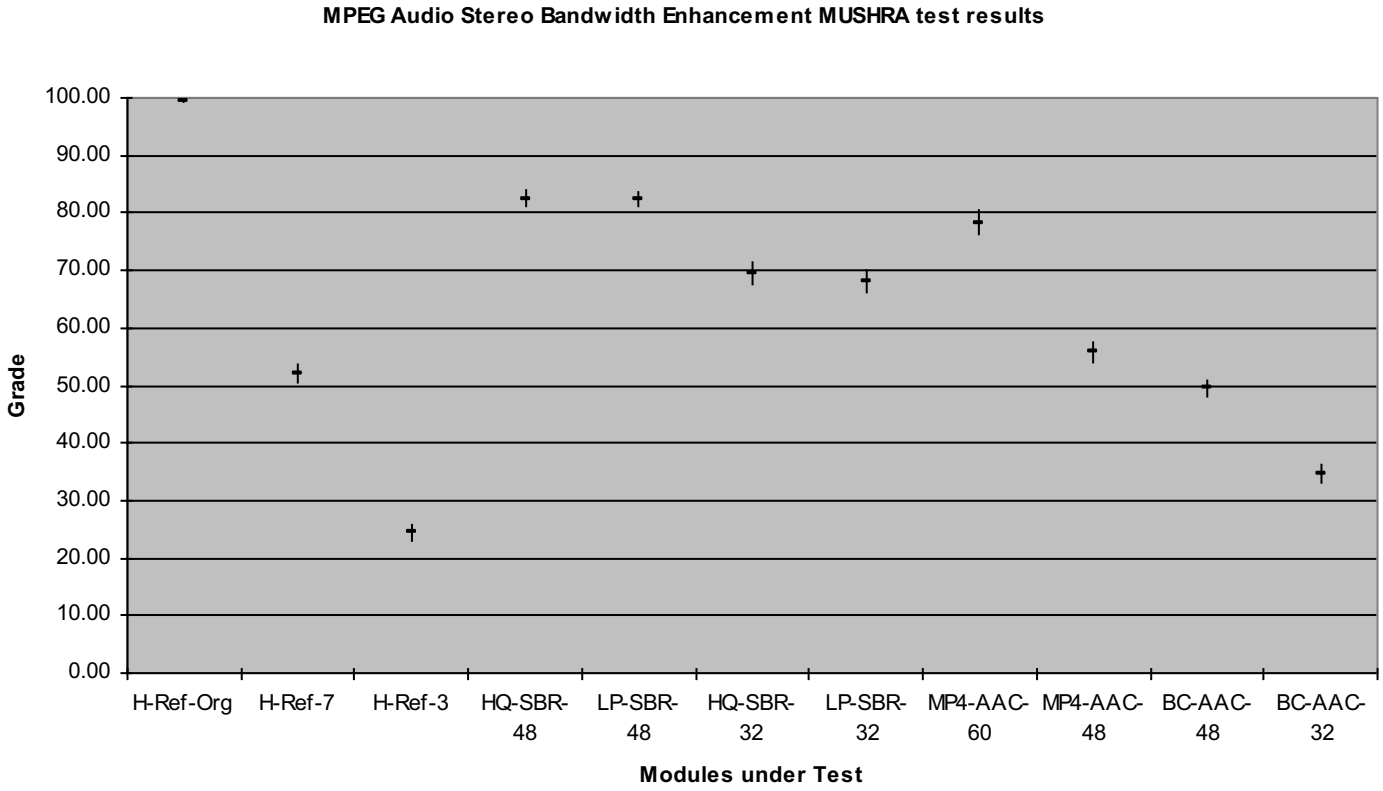

Figure 2- Stereo Test Results. See Table 1 "Label" column for key to horizontal axis labels.

ANNEX A, Test results on an item per item basis

Table A.1 - Mono Test Results for each, and all excerpts. Lower and Upper indicate 95% Confidence Interval.

Excerpt	Module	Mean	Lower	Upper
All excerpts	HQ-SBR-24	75,15	73,24	77,05
	LP-SBR-24	73,12	71,08	75,16
	MP4-AAC-30	52,45	50,19	54,71
	MP4-AAC-24	45,58	43,80	47,35
	BC-AAC-24	39,37	37,61	41,13
Bagpipes	HQ-SBR-24	77,30	70,87	83,73
	LP-SBR-24	75,70	69,51	81,89
	MP4-AAC-30	55,15	49,75	60,55
	MP4-AAC-24	48,55	42,49	54,61
	BC-AAC-24	48,25	43,16	53,34
Castanets	HQ-SBR-24	63,05	56,91	69,19
	LP-SBR-24	62,80	56,38	69,22
	MP4-AAC-30	41,05	33,92	48,18
	MP4-AAC-24	31,65	26,60	36,70
	BC-AAC-24	35,35	30,50	40,20
English speech	HQ-SBR-24	67,70	62,92	72,48
	LP-SBR-24	69,15	63,72	74,58
	MP4-AAC-30	53,40	45,76	61,04
	MP4-AAC-24	47,50	42,58	52,42
	BC-AAC-24	38,65	34,54	42,76
German speech	HQ-SBR-24	59,70	52,68	66,72
	LP-SBR-24	58,00	52,05	63,95
	MP4-AAC-30	35,40	28,51	42,29
	MP4-AAC-24	44,85	39,77	49,93
	BC-AAC-24	31,80	27,12	36,48
Glockenspiel	HQ-SBR-24	79,70	74,80	84,60
	LP-SBR-24	77,70	72,94	82,46
	MP4-AAC-30	67,70	61,29	74,11
	MP4-AAC-24	43,60	36,42	50,78
	BC-AAC-24	40,70	33,65	47,75
Harpsichord	HQ-SBR-24	79,90	74,31	85,49
	LP-SBR-24	79,40	73,82	84,98
	MP4-AAC-30	40,40	32,94	47,86
	MP4-AAC-24	43,90	39,21	48,59
	BC-AAC-24	35,90	30,27	41,53
Orchestral piece	HQ-SBR-24	81,40	76,09	86,71
	LP-SBR-24	82,70	77,91	87,49
	MP4-AAC-30	50,90	44,46	57,34
	MP4-AAC-24	44,30	39,99	48,61
	BC-AAC-24	37,75	32,00	43,50
Pitch pipe	HQ-SBR-24	74,75	69,20	80,30

Excerpt	Module	Mean	Lower	Upper
	LP-SBR-24	67,20	58,72	75,68
	MP4-AAC-30	51,30	44,78	57,82
	MP4-AAC-24	45,20	41,49	48,91
	BC-AAC-24	37,05	33,21	40,89
Plucked strings	HQ-SBR-24	81,75	77,23	86,27
	LP-SBR-24	79,85	74,92	84,78
	MP4-AAC-30	57,00	50,17	63,83
	MP4-AAC-24	47,15	41,91	52,39
	BC-AAC-24	38,75	33,72	43,78
Pop music	HQ-SBR-24	81,60	75,79	87,41
	LP-SBR-24	82,00	76,75	87,25
	MP4-AAC-30	50,30	43,51	57,09
	MP4-AAC-24	44,25	37,55	50,95
	BC-AAC-24	36,45	30,48	42,42
Suzan Vega	HQ-SBR-24	68,85	61,76	75,94
	LP-SBR-24	63,00	53,85	72,15
	MP4-AAC-30	52,15	47,54	56,76
	MP4-AAC-24	45,10	40,45	49,75
	BC-AAC-24	35,30	29,28	41,32
Trumpet solo and orchestra	HQ-SBR-24	86,05	81,30	90,80
	LP-SBR-24	79,95	74,12	85,78
	MP4-AAC-30	74,65	70,15	79,15
	MP4-AAC-24	60,85	52,30	69,40
	BC-AAC-24	56,50	49,13	63,87

Table A.2 - Stereo Test Results for each, and all excerpts. Lower and Upper indicate 95% Confidence Interval.

Excerpt	Module	Average	Low	High
All	HQ-SBR-48	82,34	80,84	83,84
	LP-SBR-48	82,20	80,68	83,73
	HQ-SBR-32	69,43	67,29	71,58
	LP-SBR-32	68,04	65,93	70,15
	MP4-AAC-60	78,21	76,06	80,35
	MP4-AAC-48	55,60	53,70	57,49
	BC-AAC-48	49,33	47,64	51,01
	BC-AAC-32	34,49	32,90	36,09
Bagpipes	HQ-SBR-48	82,70	77,37	88,03
	HQ-SBR-32	63,75	56,23	71,27
	LP-SBR-48	84,90	80,60	89,20
	LP-SBR-32	62,40	54,68	70,12
	MP4-AAC-60	75,35	68,19	82,51
	MP4-AAC-48	49,25	43,91	54,59
	BC-AAC-48	45,20	41,17	49,23
	BC-AAC-32	35,30	30,29	40,31
Castanets	HQ-SBR-48	68,25	62,15	74,35
	HQ-SBR-32	57,20	49,03	65,37
	LP-SBR-48	69,25	62,86	75,64
	LP-SBR-32	58,75	51,27	66,23
	MP4-AAC-60	61,95	55,53	68,37
	MP4-AAC-48	49,85	45,03	54,67
	BC-AAC-48	38,30	33,58	43,02
	BC-AAC-32	27,40	23,16	31,64
English speech	HQ-SBR-48	87,60	83,69	91,51
	HQ-SBR-32	73,15	67,65	78,65
	LP-SBR-48	85,60	81,64	89,56
	LP-SBR-32	70,40	65,70	75,10
	MP4-AAC-60	87,80	83,64	91,96
	MP4-AAC-48	62,85	58,65	67,05
	BC-AAC-48	51,80	47,03	56,57
	BC-AAC-32	37,85	33,10	42,60
German speech	HQ-SBR-48	80,45	75,28	85,62
	HQ-SBR-32	65,60	57,71	73,49
	LP-SBR-48	80,25	75,49	85,01
	LP-SBR-32	59,75	52,50	67,00
	MP4-AAC-60	86,25	82,01	90,49
	MP4-AAC-48	59,45	53,64	65,26
	BC-AAC-48	53,30	47,91	58,69
	BC-AAC-32	37,20	31,87	42,53
Glockenspiel	HQ-SBR-48	79,45	73,85	85,05
	HQ-SBR-32	71,50	65,31	77,69
	LP-SBR-48	78,10	72,84	83,36
	LP-SBR-32	66,65	60,06	73,24
	MP4-AAC-60	78,30	72,82	83,78
	MP4-AAC-48	66,05	58,97	73,13
	BC-AAC-48	55,00	47,90	62,10

Excerpt	Module	Average	Low	High
	BC-AAC-32	34,85	29,75	39,95
Harpsichord	HQ-SBR-48	83,05	78,37	87,73
	HQ-SBR-32	63,50	54,94	72,06
	LP-SBR-48	82,50	77,02	87,98
	LP-SBR-32	61,75	52,70	70,80
	MP4-AAC-60	53,95	46,45	61,45
	MP4-AAC-48	45,70	38,53	52,87
	BC-AAC-48	44,60	39,70	49,50
	BC-AAC-32	25,25	21,18	29,32
Orchestral piece	HQ-SBR-48	87,80	84,09	91,51
	HQ-SBR-32	72,95	65,68	80,22
	LP-SBR-48	86,50	81,50	91,50
	LP-SBR-32	73,80	67,18	80,42
	MP4-AAC-60	93,10	90,65	95,55
	MP4-AAC-48	56,25	50,77	61,73
	BC-AAC-48	51,35	45,14	57,56
	BC-AAC-32	33,65	29,03	38,27
Pitch pipe	HQ-SBR-48	78,00	73,42	82,58
	HQ-SBR-32	69,70	63,00	76,40
	LP-SBR-48	77,00	72,49	81,51
	LP-SBR-32	67,50	60,02	74,98
	MP4-AAC-60	65,30	58,67	71,93
	MP4-AAC-48	43,25	37,30	49,20
	BC-AAC-48	47,95	42,95	52,95
	BC-AAC-32	33,65	28,09	39,21
Plucked strings	HQ-SBR-48	85,25	80,98	89,52
	HQ-SBR-32	70,60	63,99	77,21
	LP-SBR-48	86,20	82,14	90,26
	LP-SBR-32	70,65	64,99	76,31
	MP4-AAC-60	78,20	72,41	83,99
	MP4-AAC-48	51,65	46,80	56,50
	BC-AAC-48	48,35	44,19	52,51
	BC-AAC-32	31,05	26,92	35,18
Pop music	HQ-SBR-48	87,00	82,86	91,14
	HQ-SBR-32	80,60	75,29	85,91
	LP-SBR-48	87,95	83,17	92,73
	LP-SBR-32	82,35	77,09	87,61
	MP4-AAC-60	84,15	79,34	88,96
	MP4-AAC-48	53,70	50,53	56,87
	BC-AAC-48	45,85	40,82	50,88
	BC-AAC-32	34,45	29,94	38,96
Suzan Vega	HQ-SBR-48	84,75	79,97	89,53
	HQ-SBR-32	72,90	66,47	79,33
	LP-SBR-48	82,60	77,98	87,22
	LP-SBR-32	73,20	67,25	79,15
	MP4-AAC-60	82,60	77,52	87,68
	MP4-AAC-48	57,80	52,55	63,05
	BC-AAC-48	48,70	45,10	52,30
	BC-AAC-32	35,50	30,76	40,24
Trumpet solo and orchestra	HQ-SBR-48	83,75	79,18	88,32
	HQ-SBR-32	71,75	63,14	80,36

Excerpt	Module	Average	Low	High
	LP-SBR-48	85,60	80,46	90,74
	LP-SBR-32	69,30	61,51	77,09
	MP4-AAC-60	91,55	87,55	95,55
	MP4-AAC-48	71,35	64,06	78,64
	BC-AAC-48	61,50	53,29	69,71
	BC-AAC-32	47,75	39,65	55,85