

地上デジタル放送方式高度化に関わる
適用技術検討作業 中間報告

VVC 主観評価実験報告

2021年10月7日

デジタル放送システム開発部会
一般社団法人 電波産業会

まえがき

総務省からの諮問第 2044 号「放送システムに関する技術的条件」(2019 年 6 月 18 日)を受け、情報通信審議会放送システム委員会に地上デジタル放送方式高度化作業班が設置され、技術的条件の検討が始まった。本活動の一環として、2020 年 6 月 22 日に、ARIB に対して映像符号化及び音声符号化方式の高度化に必要な技術的検討の依頼があった。

本依頼を受け、映像符号化方式作業班は、映像符号化方式の検討を開始している。2021 年 1 月には、VVC 規格を用いた際の所要ビットレートを求めるための主観評価実験の計画案を作成し、地上デジタル放送方式高度化作業班にて報告した。

本報告書は、上記計画案に従い実施した VVC 規格の主観評価実験の結果報告である。

内容

1. 目的.....	4
2. 実験計画.....	4
2.1. 実験会場、日時.....	4
2.2. 評価映像.....	4
2.3. 符号化条件.....	6
2.4. 評価実験方法.....	7
2.5. 所要ビットレート推定の基準.....	7
3. 符号化結果.....	8
3.1. エンコーダ A.....	8
3.2. エンコーダ B.....	9
4. 実験結果.....	11
4.1. 評価者のスクリーニング.....	11
4.2. ビットレートと MOS 値との関係.....	13
4.2.1. 1080/60/P, SDR.....	13
4.2.2. 2160/60/P, SDR.....	14
4.2.3. 2160/60/P, HDR.....	14
5. 所要ビットレート.....	15
6. 考察.....	16
6.1. エンコーダ A の妥当性.....	16
6.2. 所要ビットレート推定(2160/60/P, SDR).....	17
6.3. VVC 規格の公称性能と所要ビットレートとの関係.....	17
6.4. 符号化品質改善の可能性.....	18
6.4.1. 1080/60/P, SDR.....	18
6.4.2. 2160/60/P, SDR.....	18
7. (補足)評価映像の選定.....	20
7.1. 選定基準.....	20
7.1.1. 符号化難易度の導出方法.....	20
7.2. 符号化難易度分布と選定した評価映像.....	20
7.2.1. 1080/60/P, SDR.....	20
7.2.2. 2160/60/P, SDR.....	21
7.2.3. 2160/60/P, HDR.....	21

1. 目的

VVC 規格を適用した地上波デジタル放送の運用ガイドラインや技術基準の策定に寄与することを目的とし、UHDTV 及び HDTV の VVC 符号化映像の主観評価実験を実施して、所要ビットレートを明らかにする。所要ビットレートとは、放送されるほぼ全ての映像で一定水準以上の画質を確保可能なビットレートである。

2. 実験計画

2.1. 実験会場、日時

日本放送協会放送技術研究所(世田谷区砧)

2021年5月17日～28日

2.2. 評価映像

表 2-1 に示す、映像情報メディア学会の標準動画像を評価映像として用いた。

幅広い符号化難易度分布を持ち、かつ多様な絵柄を含む評価映像群を選定するため、絵柄の重複がなく様々な絵柄が含まれること、超高精細映像の評価映像として適切であること、符号化難易度の分布が従来の評価映像と類似することを条件とした。

表 2-1 評価映像

映像形式	準拠規格	シーン
1080/60/P, SDR	Rec. ITU-R BT.709	ハイビジョン・システム評価用標準動画像第二版 B シリーズの 8 映像(図 2-1)
2160/60/P, SDR	Rec. ITU-R BT.2020	超高精細・広色域標準動画像 A シリーズ、及び B シリーズの 8 映像(図 2-2)
2160/60/P, HDR	Rec. ITU-R BT.2100	超高精細・広色域 HDR 版標準動画像 C シリーズの 8 映像(図 2-3)
2160/60/P, SDR	Rec. ITU-R BT.2020	2160/60/P, HDR 評価映像を SDR 変換したもの 主観評価実験には用いていない

評価映像は、表 2-2 に示す映像形式にて実験で使用した。

表 2-2 映像形式

項目	値
映像信号形式	Y'CbCr 4:2:2
画素ビット数	10-bit
シーン長	10 秒 900 フレーム(15 秒)の内、第 180 フレームから第 779 フレームまでを使用



<u>S201</u> Ginkgo trees	S202 Truck train	<u>S204</u> Red leaves (pan up)	<u>S209</u> Fountain (dolly)
			
S210 Studio concert	S214 Basketball	S218 Horse racing (dirt)	<u>S265</u> Fountain (chromakey)

図 2-1 1080/60/P, SDR の評価映像

			
A03 Trains C	<u>A05</u> Steel plant	A06 Festival	B06 Paddock
			
<u>B07</u> Marathon (start)	<u>B09</u> Marathon (panning)	<u>B11</u> Water polo (Scrolling)	B13 Drama (coffee)

図 2-2 2160/60/P, SDR の評価映像

			
C01 Fireworks (willow)	C05 Fireworks (barrage)	C06 Drama (standing up)	C08 Drama (sunset)
			
C11 Swim race (backstroke)	C12 Volleyball (fixed)	C15 Paddock (fixed)	C17 Horse race (homestretch)

図 2-3 2160/60/P, HDR の評価映像

2.3. 符号化条件

表 2-3 に示すエンコーダ A 及びエンコーダ B を用いて VVC 符号化を行った。エンコーダ A は、2025 年頃にハードウェアで実現可能なリアルタイムエンコーダの画質を実現するソフトウェアエンコーダである。エンコーダ B は、実装方法の違いによる符号化性能差の検証のために補助的に用いるものであり、一部のビットレートのみで使用した。実験では、エンコーダ A の符号化映像とエンコーダ B の符号化映像とを混ぜて評定者に提示した。この際、エンコーダ B の符号化映像は、符号化難易度が比較的高いもの(図 2-1 (1080/60/P, SDR) 中及び図 2-2 (2160/60/P, SDR) それぞれの中の、下線付きの 4 評価映像)のみを使用した。

また、1080/60/P, SDR の評価映像 S265 (Fountain (chromakey)) に対してエンコーダ A の符号化制御のチューニングを行った映像が提供された。このチューニング効果を確認するビューイングを、本実験とは別に実施した。

表 2-3 VVC エンコーダ

エンコーダ	説明	符号化ビットレート
A	VVC エンコーダエミュータ 総務省の技術試験事務にて試作した、2025 年頃にハードウェアで実現可能なリアルタイムエンコーダの画質を実現するソフトウェアシミュレータ	表 2-4 に記載した全ビットレートで符号化
B	Fraunhofer HHI 研究所 VVC ソフトウェアエンコーダ VVenC (Version 0.2.1.0) VVC 標準化作業で開発された、最高性能を達成する参照ソフトウェア VTM (VVC Test Model) の最適化版リアルタイム動作よりも高性能化を主眼としており、将来の放送サービスにそのまま適用することは難しい	1080/60/P, SDR は 3 Mbps 及び 7 Mbps にて符号化 2160/60/P, SDR は 10 Mbps 及び 20 Mbps にて符号

表 2-4 に示す条件にて VVC 符号化を行った。

ビットレートは、高度広帯域衛星デジタル放送の映像符号化方式検討時¹の値(HEVC を用い、1080/60/P, 2160/60/P でそれぞれ 10 Mbps – 15 Mbps, 30 Mbps – 40 Mbps で符号化)を参考に決めた。高ビットレートレンジでは VVC 方式の適用により 30% のビットレート削減が確実に見通せることから、高ビットレート側の値をそれぞれ 10 Mbps ($\approx 15 \times 0.7$), 30 Mbps ($\approx 40 \times 0.7$) とした。低ビットレートレンジの値は、これらの値の 70%, 50%, 30% とした。

表 2-4 符号化条件

項目	値	
プロファイル	Main 10 (10-bit, 4:2:0)	
ビットレート	1080/60/P 2160/60/P	3 Mbps, 5 Mbps, 7 Mbps, 10 Mbps 10 Mbps, 15 Mbps, 20 Mbps, 30 Mbps
符号化パラメータ	GOP 構造	階層 B 参照 (ARIB STD-B32 に記載の L3 構造)
	IRAP 間隔	32/60 sec
	GOP 長	8 フレーム
	CPB サイズ	1 秒分

¹ 実験報告書は https://www.soumu.go.jp/main_content/000262094.pdf である。

項目	値	
	色差信号 サンプリング位置	垂直・水平方向共、輝度信号位置と同じ (vui_chroma_sample_loc_type_frame = 2)
ツール設定	エンコーダ A	スクリーンコンテンツ向けツール(IBC)は無効化 他のツールは非公開
	エンコーダ B	Medium 設定、MCTF (プレフィルタ)は無効化

2.4. 評価実験方法

表 2-5 に示す方法にて実施した。

表 2-5 評価実験方法

項目	内容
評価方法	二重刺激劣化尺度(DSIS)法、5段階劣化尺度(表 2-6) 基準映像—評価映像のペアを一回提示
観視条件	Rec. ITU-R BT.500-14
評価者	専門家 38 人 注記: 装置故障により、1 人の 2160/60/P, SDR の実験データが欠損
ディスプレイ	PVM-X550 (55-inch LCD)
視距離	画面高の 1.5 倍(1.5H)(2160/60/P 時) 画面高の 3 倍(3H)(1080/60/P 時)

表 2-6 5段階劣化尺度

評点	評価語
5	劣化が分からない
4	劣化が分かるが気にならない
3	劣化が気になるが邪魔にならない
2	劣化が邪魔になる
1	劣化が非常に邪魔になる

一回の実験時間は、約 2 時間 30 分となった。内訳は、3 つの映像フォーマットそれぞれの観視・評価時間 (20 分×3)、説明時間(30 分)、及び休憩時間(30 分×2)である。なお、新型コロナウイルス対策のため、一日当たりの評価者数上限を 6 名(午前 3 名、午後 3 名)とした。

2.5. 所要ビットレート推定の基準

放送品質を満足するサービスを提供するためには、画質の許容限界と見なされる平均評価値(Mean Opinion Score, MOS) 3.5 以上がほぼ全ての映像で満足することが望まれ、また MOS 3 未満は放送品質として許容しがたいと考えられる。そこで、平均値と分散を元に、MOS 3.5 以上及び MOS 3 未満に該当するか否かを、有意水準 5%で検定する。

3. 符号化結果

各映像フォーマットの客観画質(15秒平均)を以下に示す。

3.1. エンコーダ A

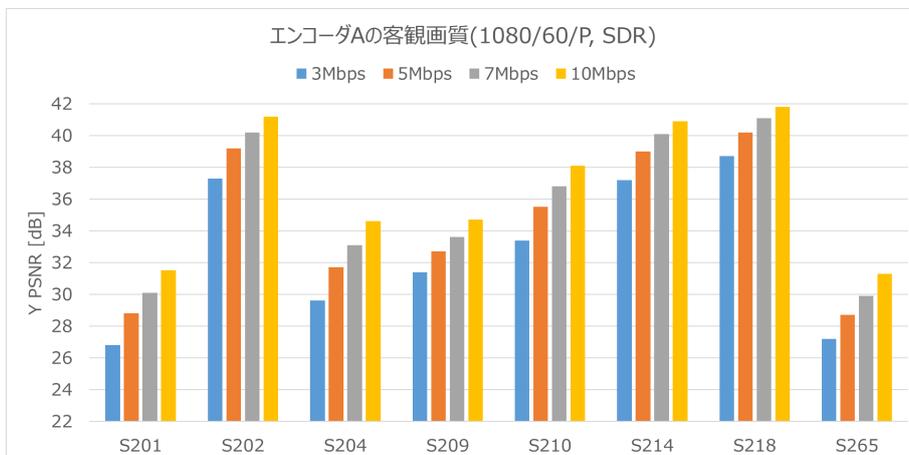


図 3-1 エンコーダ A の客観画質(1080/60/P, SDR)

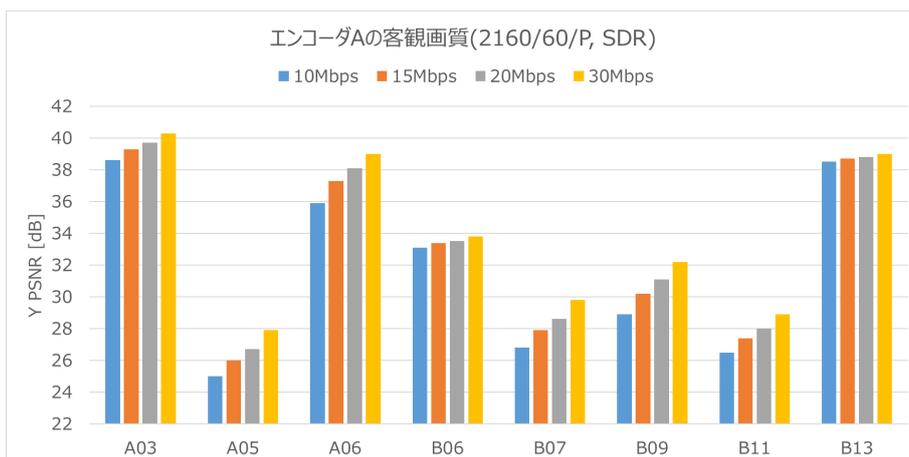


図 3-2 エンコーダ A の客観画質(2160/60/P, SDR)

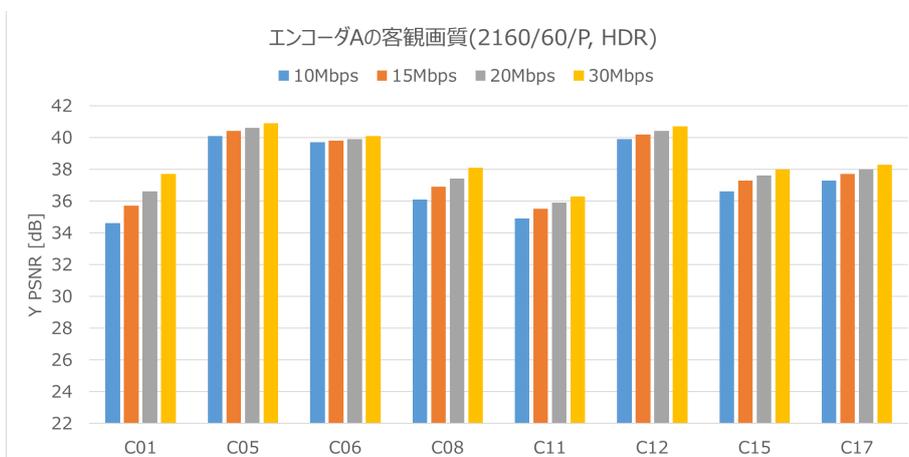


図 3-3 エンコーダ A の客観画質(2160/60/P, HDR)

3.2. エンコーダ B

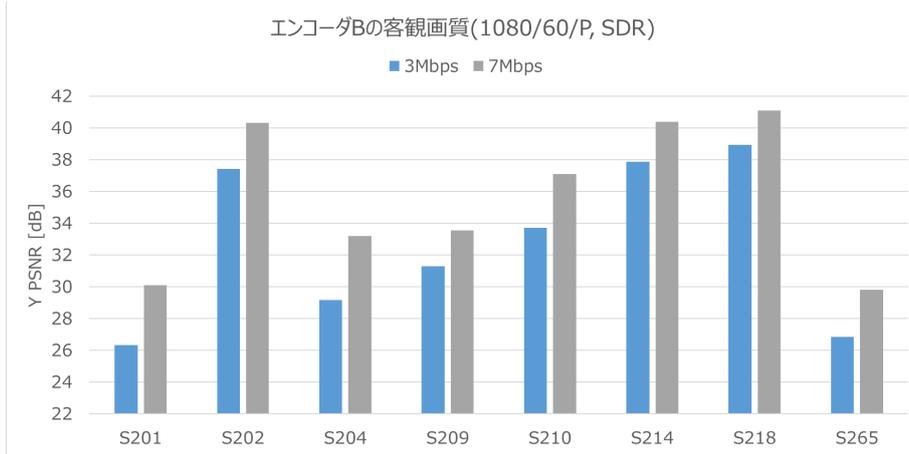


図 3-4 エンコーダ B の客観画質(1080/60/P, SDR)

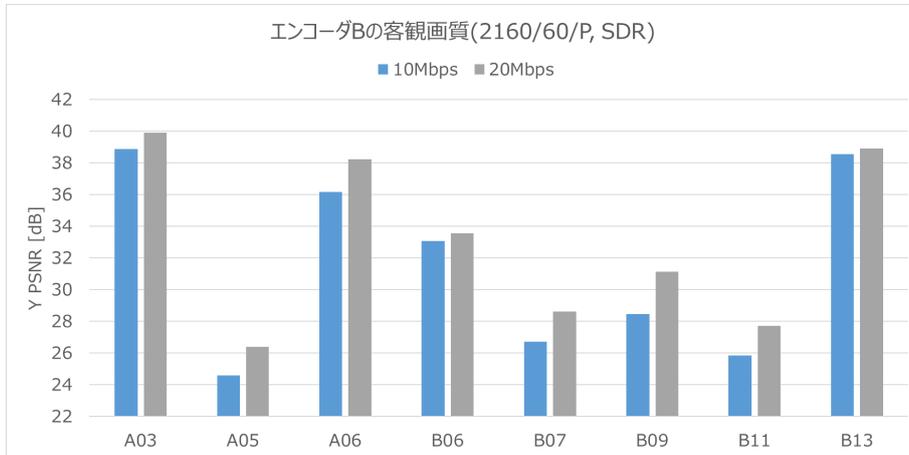


図 3-5 エンコーダ B の客観画質(2160/60/P, SDR)

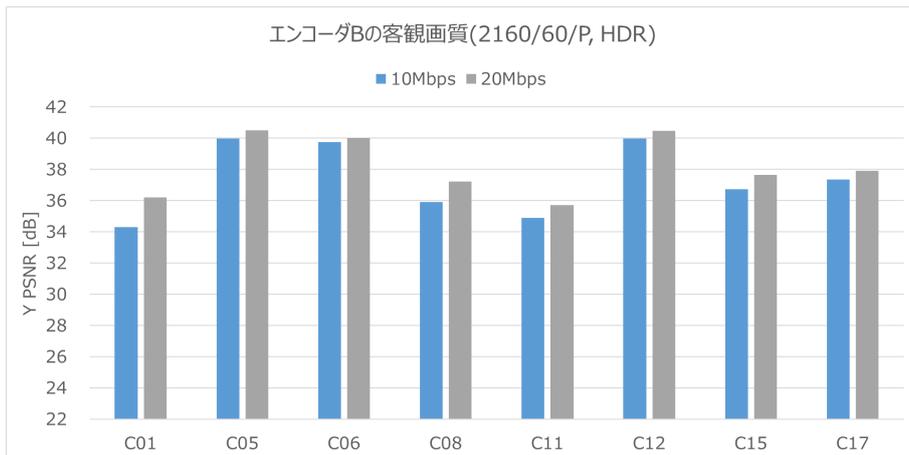


図 3-6 エンコーダ B の客観画質(2160/60/P, HDR)

2160/60/P, HDR 映像を SDR 変換した映像の客観画質を図 3-7、2160/60/P, HDR 映像と SDR 変換した映像との比較を図 3-8 に示す。SDR 変換により概ね 3 dB 低下する。

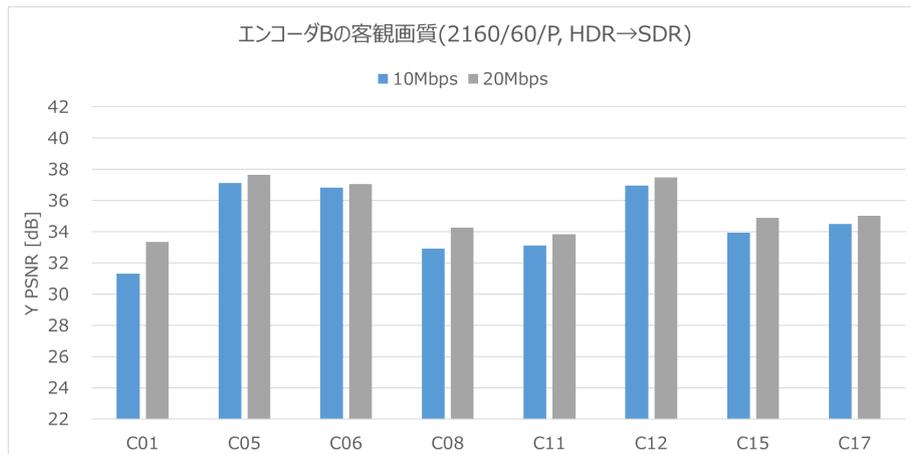


図 3-7 エンコーダ B の客観画質(2160/60/P, HDR→SDR)

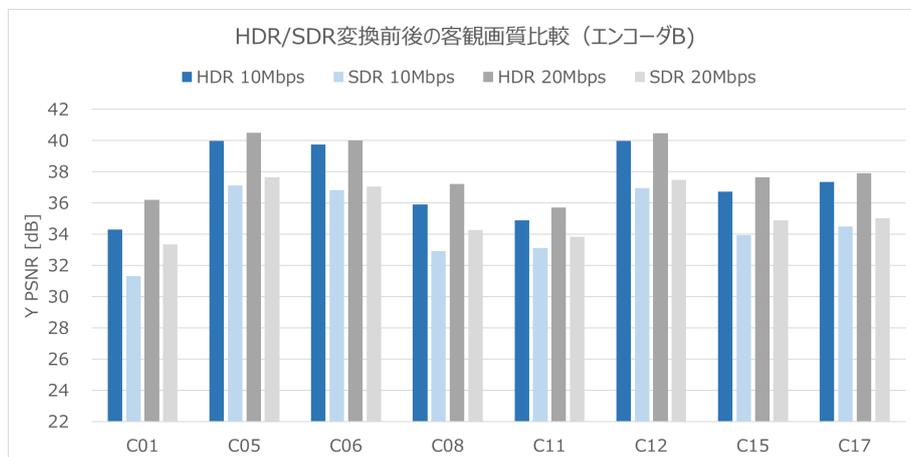


図 3-8 エンコーダ B における 2160/60/P, HDR 映像を SDR 変換した場合の客観画質の違い

4. 実験結果

4.1. 評価者のスクリーニング

3つの映像フォーマットそれぞれで、評価者のスクリーニングを行った。外れ値によって除外される評価者はいなかった。

更に、Pearson 相関に基づくスクリーニングを行った。図 4-1、図 4-2、図 4-3 に各映像フォーマットにおける、各評価者の評価結果と評価者平均値との Pearson 相関を示す。2160/60/P, HDR のみ、Pearson 相関が 0.4 未満の評価者が 7 名おり、これら評価者の評価結果を除外した。スクリーニング後の評価者数は 1080/60/P, SDR、2160/60/P, SDR、2160/60/P, HDR でそれぞれ、38 名、37 名、31 名となった。

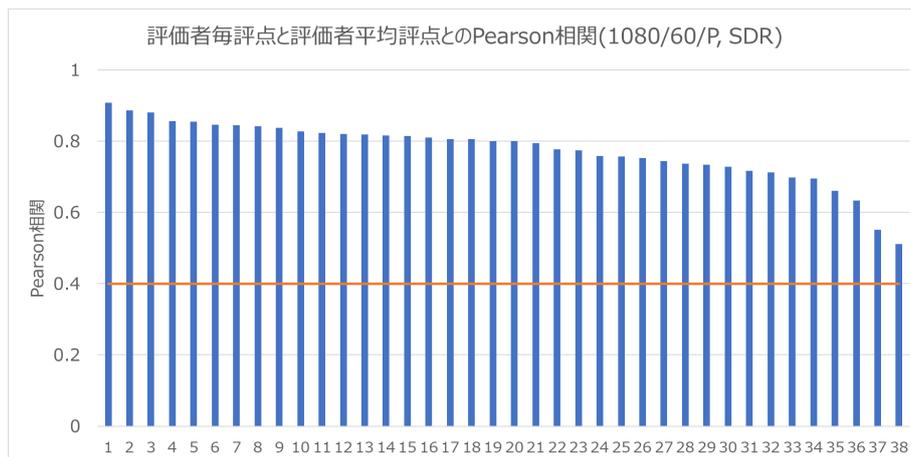


図 4-1 評価者毎評点と評価者平均評点との Pearson 分布(1080/60/P, SDR)

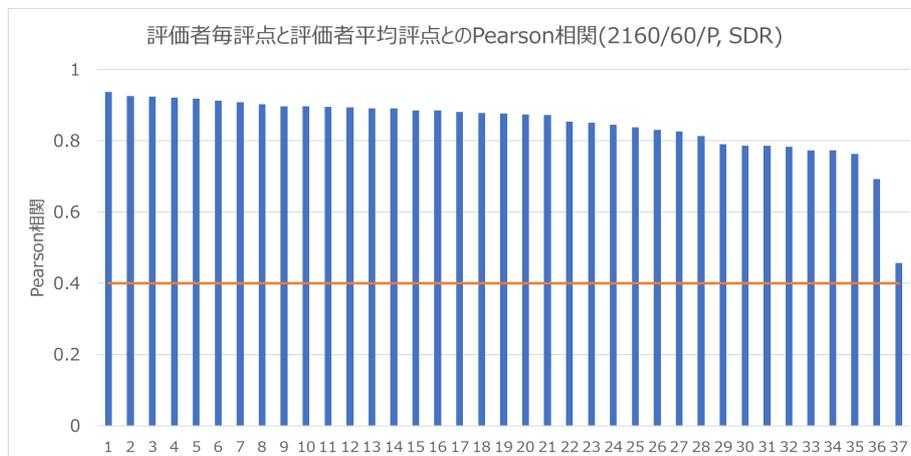


図 4-2 評価者毎評点と評価者平均評点との Pearson 分布(2160/60/P, SDR)

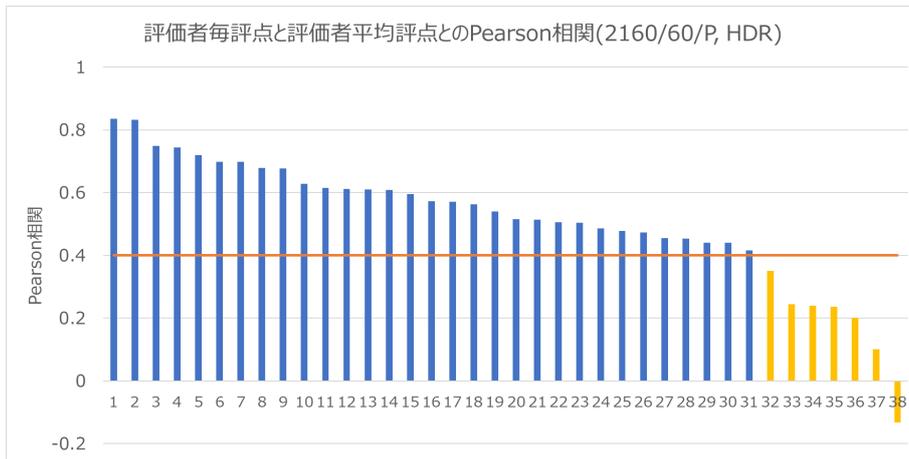


図 4-3 評価者毎評点と評価者平均評点との Pearson 分布(2160/60/P, HDR)

4.2. ビットレートと MOS 値との関係

エンコーダ A の符号化映像について、各映像フォーマットでのビットレートと MOS 値との関係、及び画質基準に対する評価映像の割合を示す。表の「MOS 値 3.5 以上」及び「MOS 値 3.0 未満」の算出には、エラーオーバー上限値を用いている。また、参考値である「MOS 値平均」の算出には、平均値を用いている。

4.2.1. 1080/60/P, SDR

最も符号化難易度が高い評価映像 S265 の MOS 値は、ビットレートが 7 Mbps 以上の時に 3.0 以上となる。また、5 Mbps 以上の時に、S265 以外の全評価映像の MOS 値が 3.5 以上となる。

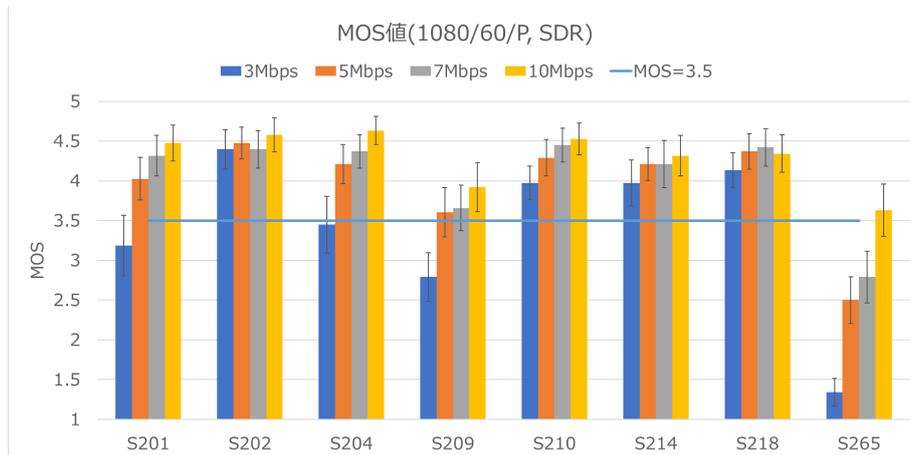


図 4-4 1080/60/P, SDR の評価映像におけるビットレートと MOS 値との関係(平均値と 95%信頼区間)

表 4-1 1080/60/P, SDR の評価映像におけるビットレートと画質基準に対する評価映像の割合

	3 Mbps	5 Mbps	7 Mbps	10 Mbps
MOS 値 3.5 以上	6 / 8	7 / 8	7 / 8	8 / 8
MOS 値 3.0 未満	1 / 8	1 / 8	0 / 8	0 / 8
MOS 値平均(参考)	3.4	4.0	4.1	4.3

4.2.2. 2160/60/P, SDR

最も符号化難易度が高い評価映像 A05 の MOS 値は、ビットレートが 30 Mbps 以上の時に 3.0 以上となる。また、30 Mbps 以上の時に、A05 以外の全評価映像の MOS 値が 3.5 以上となる。

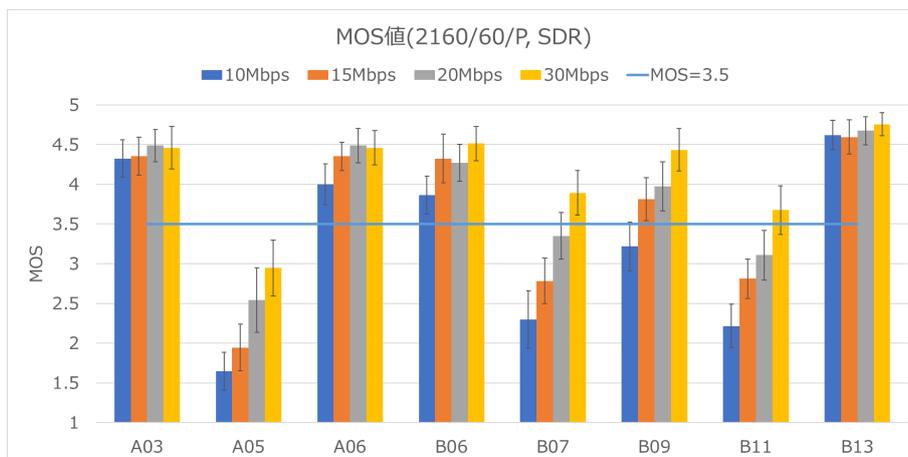


図 4-5 2160/60/P, SDR の評価映像におけるビットレートと MOS 値との関係(平均値と95%信頼区間)

表 4-2 2160/60/P, SDR の評価映像におけるビットレートと画質基準に対する評価映像の割合

	10 Mbps	15 Mbps	20 Mbps	30 Mbps
MOS 値 3.5 以上	5 / 8	5 / 8	6 / 8	7 / 8
MOS 値 3.0 未満	3 / 8	1 / 8	1 / 8	0 / 8
MOS 値平均(参考)	3.3	3.6	3.9	4.1

4.2.3. 2160/60/P, HDR

ビットレートが 10 Mbps の時に、全ての評価映像の MOS 値が 3.0 以上となる。

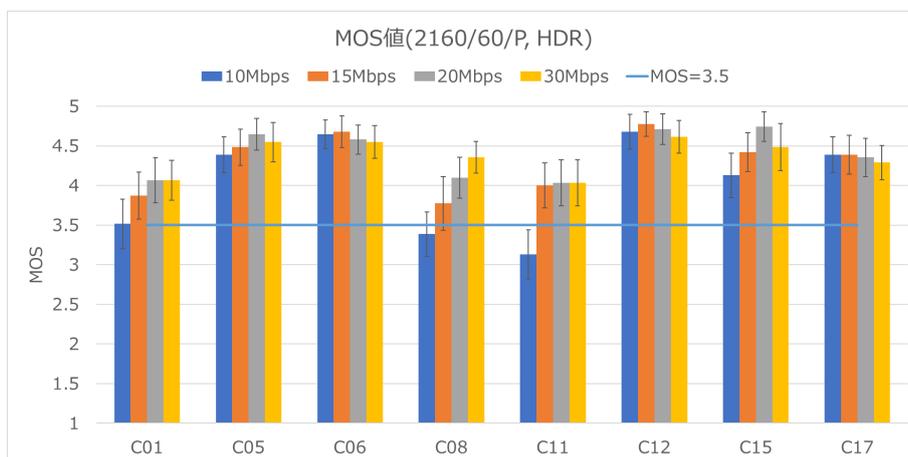


図 4-6 2160/60/P, HDR の評価映像におけるビットレートと MOS 値との関係(平均値と95%信頼区間)

表 4-3 2160/60/P, HDR の評価映像におけるビットレートと画質基準に対する評価映像の割合

	10 Mbps	15 Mbps	20 Mbps	30 Mbps
MOS 値 3.5 以上	7 / 8	8 / 8	8 / 8	8 / 8
MOS 値 3.0 未満	0 / 8	0 / 8	0 / 8	0 / 8
MOS 値平均(参考)	4.0	4.3	4.4	4.4

5. 所要ビットレート

2.5 節の判断基準から、VVC 方式適用時の所要ビットレートは表 5-1 のように導出される。なお、2160/60/P 映像の所要ビットレートは、SDR 映像と HDR 映像で分けず、SDR 映像で得られた値とした。

今回用いた HDR 映像の符号化難易度は SDR 映像に比べて低く、図 7-2、図 7-3 に示すように SDR 映像における符号化難易度の低い 4 つのシーケンス(A03, A06, B06, B13)に相当する。これらのシーケンスは SDR 映像の主観評価実験においていずれも 10 Mbps で MOS 値 3.5 以上の要求品質を満たしており(図 4-5)、HDR 映像の主観評価結果(図 4-6)と整合している。また、実際の運用では 2160/60/P, SDR 映像相当の符号化難易度を持つ 2160/60/P, HDR 映像が使用されると想定される。これらのことから、HDR 映像には SDR 映像と同等のビットレートが要求されると考えられる。

表 5-1 VVC 方式適用時の所要ビットレート

	1080/60/P	2160/60/P
所要ビットレート	7 Mbps	30 Mbps

6. 考察

6.1. エンコーダ A の妥当性

エンコーダ A の符号化映像と、エンコーダ B の符号化映像との MOS 値の違いを図 6-1 及び図 6-2 に示す。映像は、本実験の評価映像の中で符号化難易度が比較的高いものである。

これらの図の通り、エンコーダ A とエンコーダ B との MOS 値差は、シーンによって変わるものの、平均して同等と見なせる。このことから、本実験結果は、エンコーダ A に特化した結果ではないと言える。

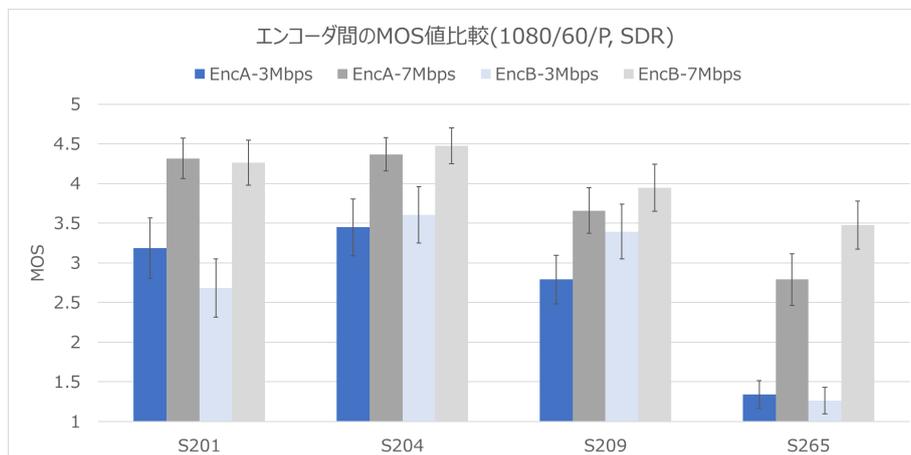


図 6-1 エンコーダ A とエンコーダ B の MOS 値比較 (1080/60/P, SDR 映像)

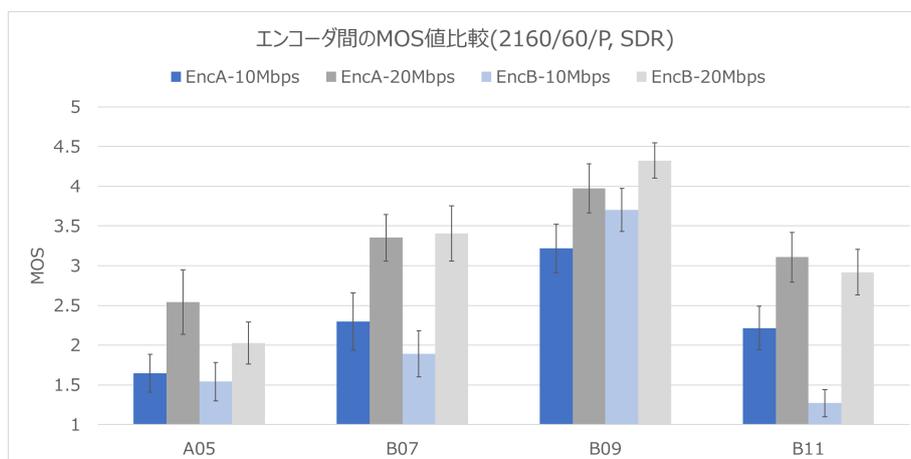


図 6-2 エンコーダ A とエンコーダ B の MOS 値比較 (2160/60/P, SDR 映像)

6.2. 所要ビットレート推定(2160/60/P, SDR)

5章では、2160/60/P, SDR の所要ビットレートを 30 Mbps とした。本実験における次に小さいビットレートは 20 Mbps であり、2.5 節の判断基準を満たすビットレートは 20 Mbps と 30 Mbps の間にあると考えられる。そこで、20 Mbps と 30 Mbps の間を直線近似し、判断基準を満足するビットレートを推定した。

図 4-5 及び表 4-2 から、評価映像 A05 の MOS 値が 3.0 以上、かつ評価映像 B11 の MOS 値が 3.5 以上となるビットレートが、推定するビットレートとなる。図 6-3 及び図 6-4 は、それぞれのビットレートと MOS 値(実線は平均値、点線はエラーバー上限及び下限)との関係である。エラーバー上限が上記の値となるビットレート 22 Mbps が、実際の所要ビットレートと推定される。

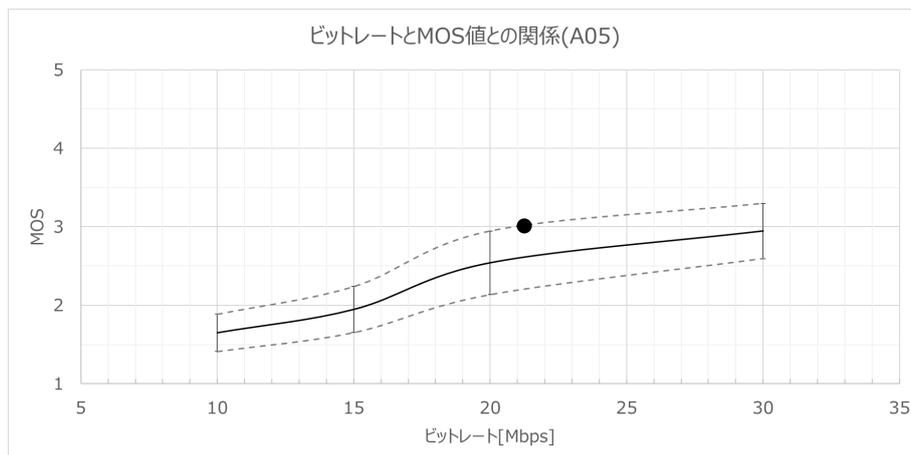


図 6-3 ビットレートと MOS 値との関係 (評価映像 A05)

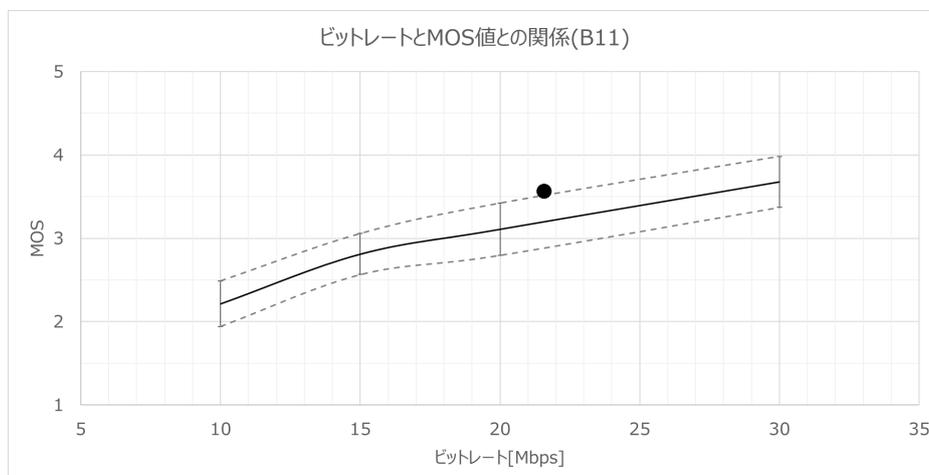


図 6-4 ビットレートと MOS 値との関係 (評価映像 B11)

6.3. VVC 規格の公称性能と所要ビットレートとの関係

VVC 規格の HEVC 規格からの圧縮効率向上は、客観画質評価に基づく分析によって 40%弱、主観画質評価に基づく分析によって 50%弱(共に SDR 映像)と、VVC 標準化を担う機関から報告されている。これらは、評価に用いられた少数の評価映像の平均的な数値である。一方、本報告における所要ビットレートは、符号化難易度の高い映像を含む評価画像のほぼ全てで一定水準以上の画質を確保可能なビットレートである。この違いに留意する必要がある。

近年の映像符号化方式は、映像に存在する冗長性を徹底的に取り除くことで圧縮効率を伸ばしている。冗長性が比較的多く存在し符号化難易度が低い映像では公称性能に近い圧縮効率向上となる一方、冗長性が少なく符号化難易度が高い映像では小さくなる。極端な例では、冗長性の無いノイズ映像は圧縮効率

の向上を期待できない。

今回の実験結果で得られた 1080/60/P 映像及び 2160/60/P 映像の VVC による所要ビットレートと高度広帯域衛星デジタル放送の映像符号化方式検討時の HEVC による所要ビットレートを表 6-1 に示す。1080/60/P では 30%以上の圧縮効率向上が認められる一方で、2160/60/P では圧縮効率の向上が小さい。

表 6-1 VVC と HEVC の所要ビットレート

注: 評価映像等が異なるため、本表は両規格の性能差を表すものではない

	VVC	HEVC
1080/60/P	7 Mbps	10 – 15 Mbps
2160/60/P	30 Mbps	30 – 40 Mbps

符号化難易度が高い映像を除いた他の 2160/60/P 映像における、MOS 値 3.5 を超えるビットレートの比較を表 6-2 に示す。HEVC の場合の 15 Mbps に対して、VVC では 10 Mbps であり、30%以上の圧縮効率向上が認められる。

表 6-2 符号化難易度が高い映像を除いた他の 2160/60/P 映像における、MOS 値が 3.5 を超えるビットレート

VVC	HEVC	HEVC からの 圧縮効率向上率
10 Mbps	15 Mbps	約 33%

6.4. 符号化品質改善の可能性

6.4.1. 1080/60/P, SDR

2.3 節のように、符号化制御チューニングを行ったエンコーダ A の符号化映像(評価映像 S265 のみ)が提供されたため、チューニング前(本実験対象)とチューニング後(本実験対象外)との主観的な画質比較を、実験参加者へのデモセッション(実験後、表 2-5 の条件に合致しない環境)にて実施した。

符号化制御チューニングを行うことによって、主観画質が概ね一つ上のビットレート(例えば、チューニング後の 5 Mbps 映像の主観画質は、チューニング前の 7 Mbps 映像のものと同様)となることが確認された。他の評価映像では未確認ではあるが、このような符号化制御チューニングは、HEVC 方式の実用化時にも同様な効果を生んだものであるため、他のシーンへの悪影響はないものと推定される。

4.2.1 節のように、所要ビットレートは評価映像 S265 の MOS 値によって 7 Mbps という結論となっているが、上記の結果を考慮すると、実用的には 5 Mbps でも基準を満足する可能性が高いものと思われる。

6.4.2. 2160/60/P, SDR

2160/60/P SDR の評価映像の中で、特に A05 の符号化難易度が高い。2160/60/P SDR の評価映像セットの中には、他にも A07 という非常に符号化難易度が高い映像がある(図 7-2 参照)。事前の主観画質確認では、本実験よりも更に高いビットレートでの符号化でないと十分な主観画質が得られないことが判明している。

VVC 規格は、従来規格にはない新たな符号化ツールとして、例えば RPR (Reference Picture Resampling)² が導入されている。RPR を用いて、符号化難易度が高いシーンでは空間解像度を落として(例えば 2160/P を 1080/P に)符号化することで、主観画質の向上が期待できる。RPR は本実験で使用したエンコーダに実

² 動き補償時に、処理対象ピクチャと参照ピクチャとの画面サイズ比を考慮したスケーリングとフィルタリングを行う技術である。これにより、各ピクチャで異なる画面サイズを取ることが可能になる。

装されていないため、使用していない。

今後、RPR など今回の実験で使用していない VVC 規格の新たな符号化ツールや、従来装置で導入されている時空間フィルタの適用で品質向上が可能か、検討を行う予定である。

7. (補足)評価映像の選定

7.1. 選定基準

高度広帯域衛星デジタル放送の映像符号化方式検討(以下、HEVC 実験)の時と同様に、多様な符号化難易度や絵柄を含む評価映像を選択した。符号化難易度は、輝度信号の PSNR とする。

7.1.1. 符号化難易度の導出方法

VVenC で符号化した際の符号化難易度とした。符号化条件を表 7-1 に示す。

表 7-1 符号化条件

項目	内容
エンコーダ	VVenC version 0.2.1.0
ツール設定	Medium 設定、MCTF は無効化
符号化パラメータ	Random access GOP 長 8 フレーム IRAP 間隔 32 フレーム
ビットレート	1080/60/P: 8 Mbps 2160/60/P: 20 Mbps
処理フレーム数	900 フレーム

7.2. 符号化難易度分布と選定した評価映像

映像情報メディア学会標準動画像における符号化難易度の分布と、選定した映像を映像フォーマット毎に示す。符号化難易度が高いシーンを中心に、中間のシーンも含めて選定している。図右側は、HEVC 実験時の 4320/60/P 評価映像を各解像度にダウンサンプルした後に、同一条件で符号化した場合の符号化難易度である(2160/60/P, HDR は除く)。

7.2.1. 1080/60/P, SDR

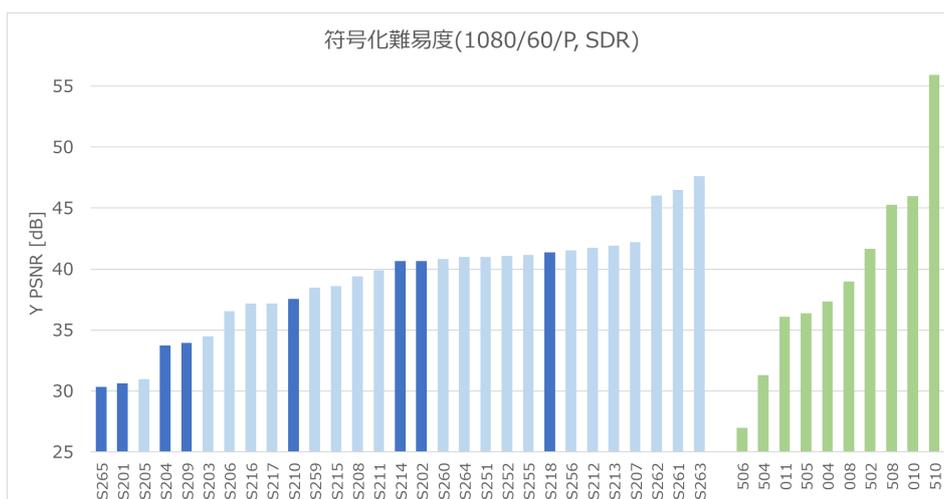


図 7-1 1080/60/P 映像の符号化難易度、選択した評価映像(濃青線で表示)

7.2.2. 2160/60/P, SDR

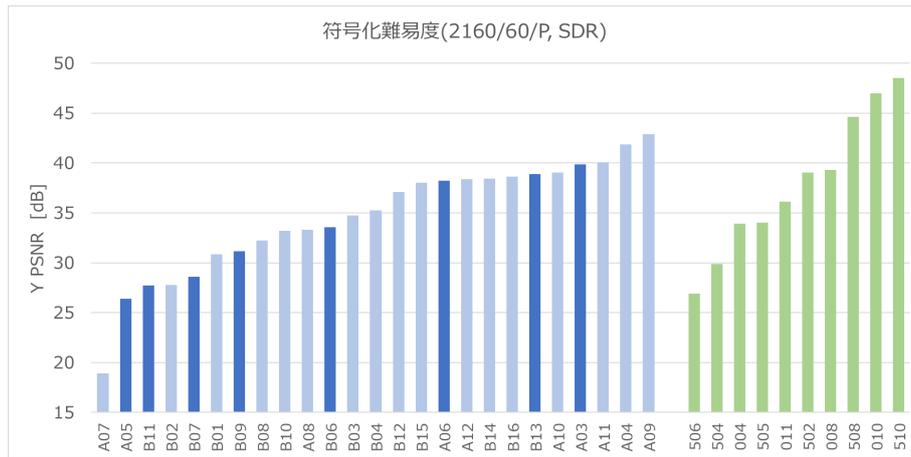


図 7-2 2160/60/P, SDR 映像の符号化難易度、選択した評価映像(濃青線で表示)

7.2.3. 2160/60/P, HDR

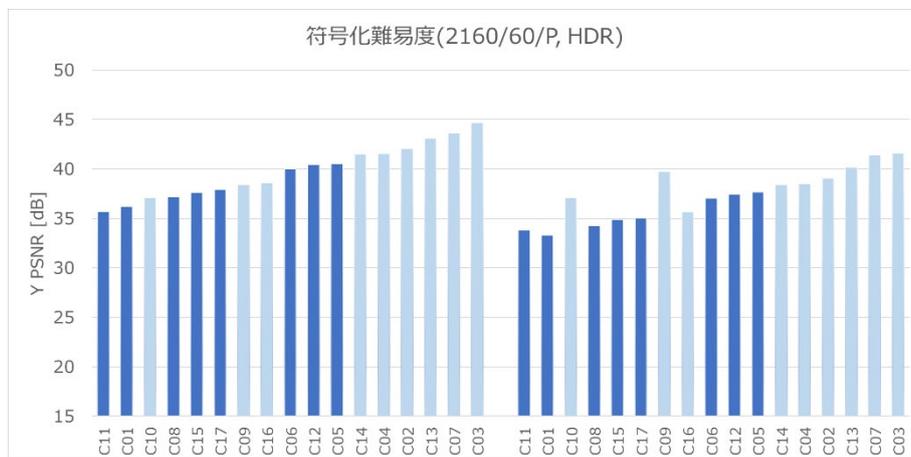


図 7-3 2160/60/P, HDR 映像の符号化難易度、選択した評価映像(濃青線で表示)
右側は SDR 変換した場合の符号化難易度