



HAL
open science

Exploring Bitrate Costs for Enhanced User Satisfaction: A Just Noticeable Difference (JND) Perspective

Hadi Amirpour, Jingwen Zhu, Raimund Schatz, Patrick Le Callet, Christian
Timmerer

► To cite this version:

Hadi Amirpour, Jingwen Zhu, Raimund Schatz, Patrick Le Callet, Christian Timmerer. Exploring Bitrate Costs for Enhanced User Satisfaction: A Just Noticeable Difference (JND) Perspective. Data Compression Conference, IEEE, Mar 2024, Snowbird, United States. hal-04497794

HAL Id: hal-04497794

<https://hal.science/hal-04497794v1>

Submitted on 11 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploring Bitrate Costs for Enhanced User Satisfaction: A Just Noticeable Difference (JND) Perspective

Hadi Amirpour^{1,*}, Jingwen Zhu^{2,*}, Raimund Schatz³,
Patrick Le Callet², and Christian Timmerer¹

¹Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria

³Nantes Université, Ecole Centrale Nantes, CNRS, LS2N, UMR 6004, Nantes, France

³AIT Austrian Institute of Technology, Austria

* These authors contributed equally to this work

Abstract

The evolving landscape of the delivery of multimedia content requires a deep understanding of how the design of the bitrate ladder (for HTTP adaptive streaming) influences cost and quality. This paper explores the use of Just Noticeable Differences (JND) to select bitrate-resolution pairs for constructing a bitrate ladder with respect to the proportion of satisfied user ratio (SUR). To expand the investigation to various codecs, first, a method is explained that transfers the JND points obtained through subjective testing from one codec (e.g., AVC) to other codecs (e.g., HEVC, VVC). This approach helps avoid the additional costs associated with conducting subjective tests to obtain JND points for a wide range of different codecs. To achieve this objective, we investigate the codec-agnostic nature of various video quality metrics, followed by the transfer of JND between two codecs, taking into account the most suitable codec-agnostic video quality metric. Secondly, we delve into the analysis of the bitrate cost of a given bitrate ladder from a JND perspective, i.e., as a function of the SUR. Among others, our experimental results demonstrate that to increase the SUR from 75% to 90%, it is necessary to double the video bitrate.

1 Introduction

The usage of video streaming platforms such as YouTube, Netflix, Hulu or Amazon Prime Video has become an integral part of our daily lives. In this context, *HTTP Adaptive Streaming* (HAS) has become the dominant technique utilized for both live and Video-on-Demand (VoD) streaming applications. HAS relies on Adaptive Bitrate (ABR) Streaming methods [1] which encode video content at multiple bitrate-resolution pairs known as "representations". These different representations are used to construct a so-called bitrate ladder [2], allowing a dynamic adjustment of video quality that takes into account the available bandwidth of the viewer and the type of device.

So far, the norm has been to utilize a fixed set of representations, such as the HLS bitrate ladder [3] for all video content on a system or platform. However, such a "one-size-fits-all" approach may not be optimal when provisioning a wider range of video content types. For this reason, *per-title encoding* approaches were introduced, which aim to create an individual, optimized bitrate ladder for each video content in order to achieve higher Quality of Experience (QoE). In per-title encoding [4, 2, 5], various encoding parameters (such as frame rate, resolution, etc.) are varied and

utilized by encoding content clips using all possible combinations of these parameters. Subsequently, an optimized bitrate ladder is constructed by selecting representations from a convex-hull based on the quality measurements of the encoded representations (cf. Fig. 1). In terms of objective quality metrics, Video Multi-method Assessment Fusion (VMAF) [6] is frequently used, due to its strong correlation with human-perceived quality (QoE). Therefore, VMAF is often used to assess the quality of representations and guide the bitrate laddering process [7].

A key step in constructing an optimized bitrate ladder is the careful selection of a subset of representations from the convex-hull. This selection process has to consider a broad range of factors, including available network bandwidth, device capabilities. While some methods focus on selecting representations based on the probability of clients requesting specific bitrate versions [8, 9], other approaches prioritize a selection of representations that minimize perceptual similarity [10, 11]. These methods aim to avoid including representations in the bitrate ladder that feature overly similar perceptual qualities, since such redundancy may lead to inefficient resource utilization. By diversifying the quality levels of the representations, the bitrate ladder construction aims to provide a wider range of viewing options and enhance the overall streaming experience for users. Figure 6 shows an example of selecting bitrate-resolution pairs from the convex-hull. In this example, selection is based on the optimal encoding parameters for each bitrate, by focusing either on bitrate or quality. Figure 1a illustrates the selection process based on bitrate, similar to the approach described in [9]. In this method, the most frequently requested set of bitrates, *i.e.*, $\{b_1, b_2, \dots, b_n\}$, is chosen from within the convex hull to construct the bitrate ladder. Figure 1b depicts a quality-based selection process, specifically for VMAF, similar to the approaches outlined in [10, 11]. These methodologies choose a set of quality values, denoted as $\{v_1, v_2, \dots, v_n\}$ with the goal of having only small, barely noticeable perceptual differences between consecutive representations. These approaches help to minimize streaming costs by avoiding the inclusion of perceptually similar quality levels in the bitrate ladder.

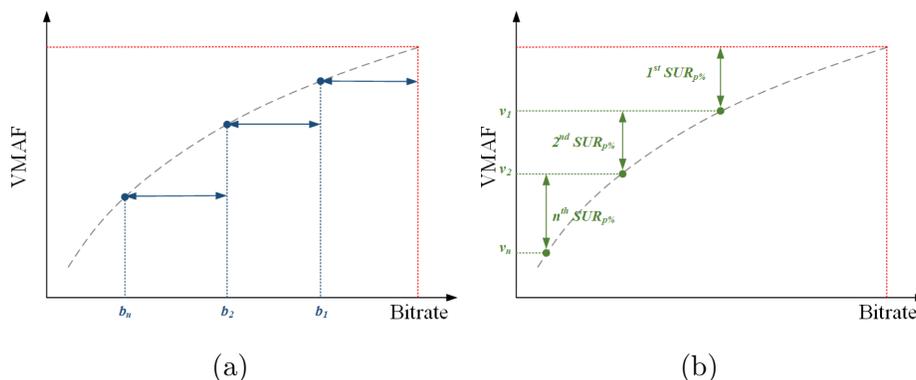


Figure 1: Optimal representation selection along the convex hull based on (a) bitrate or (b) quality (using VMAF metric).

The removal of similar quality bitrate-resolution pairs from the bitrate ladder is rooted in the understanding that the human visual system (HVS) has well-known limitations in detecting minor distortions in videos [12, 13, 14], attributed to various

psychological and physiological mechanisms. The key point is that the HVS can differentiate only a few discrete-scale distortion levels within a wide range of quality levels. The minimum visual difference that can be perceived by the HVS of *each individual* is defined as the Just Noticeable Difference (JND). In video coding, the 1st JND marks where a viewer notices quality degradation from an uncompressed reference. The 2nd JND indicates further noticeable quality drop using the 1st JND as a reference. The 3rd JND represents an additional decline in quality, with the 2nd JND as its reference. To account for variations in HVS *between viewers*, a metric called Satisfied User Ratio (SUR) was defined, representing the proportion of the population that could distinguish a given distortion level [12, 15, 16]. The 1st SUR($p\%$), for example, shows the $p\%$ of viewers who see the first JND level as good as uncompressed video.

Figure 2 presents an example of an SUR curve based on the VMAF metric. Obviously, as the number of satisfied individuals p increases (equating a decreasing audience share that is actually capable of perceiving the distortion between the reference and degraded video), the VMAF value shows a corresponding increase. This observation implies that to satisfy a larger number of individuals, the VMAF score of the encoded video should be improved, signifying that the quality of the encoded video should more closely match that of the reference video.

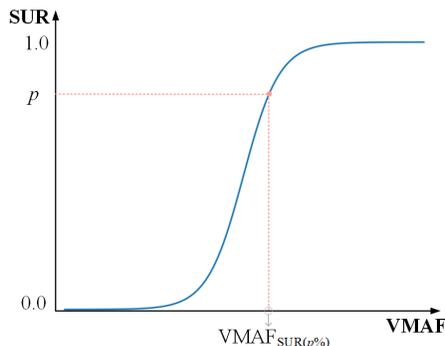


Figure 2: SUR curve example for the VMAF proxy.

For example, consider $\text{VMAF}_{\text{SUR}(75\%)}$, which signifies that 75% of viewers cannot detect any distortion between the reference video and the encoded video when the VMAF score reaches this level. To cater to a larger audience and meet higher quality standards, where 95% of viewers should not perceive any differences between the reference video and the encoded video, we would refer to the VMAF score as $\text{VMAF}_{\text{SUR}(95\%)}$. In this context, it is important to note that $\text{VMAF}_{\text{SUR}(75\%)}$ would be less than $\text{VMAF}_{\text{SUR}(95\%)}$, indicating a higher level of satisfaction in the latter case.

Figure 3 illustrates how various values of p can influence the selection of bitrate-resolution pairs. As the ratio of satisfied individuals, represented by p , increases, the chosen bitrates for the bitrate ladder also increase. In this paper, our goals are two-fold:

(i) Our objective is to quantify the relationship between the percentage of satisfied individuals p and the overall bitrate of the selected representations for the bitrate ladder:

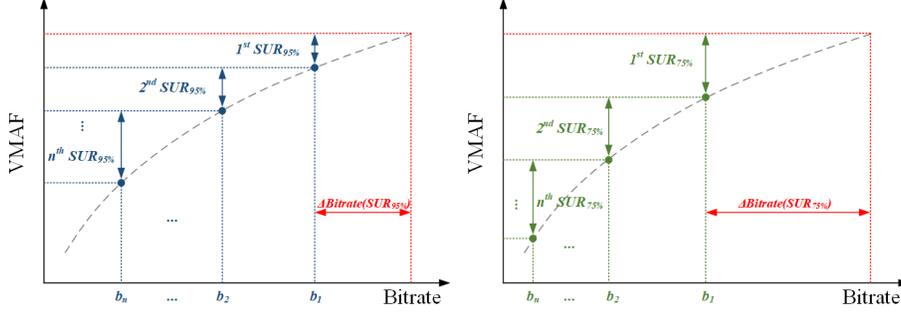


Figure 3: A growing percentage of satisfied users ($\%p$) requires a bitrate ladder with a broader range of bitrate values.

$$\sum_{i=1}^n b_i = f(p) \quad (1)$$

(ii) To accommodate different codecs, we propose a mapping scheme for SUR($p\%$) across various codecs. To achieve this, we initially assess the codec-dependent nature of different Video Quality Metrics (VQMs) by analyzing a subjective test results. Subsequently, we employ the most codec-agnostic VQM to map SUR($p\%$) to other codecs. This approach eliminates the necessity for conducting subjective tests for different video codecs. Figure 4 illustrates the mapping of the corresponding QP values for SUR($p\%$) from AVC to VVC.

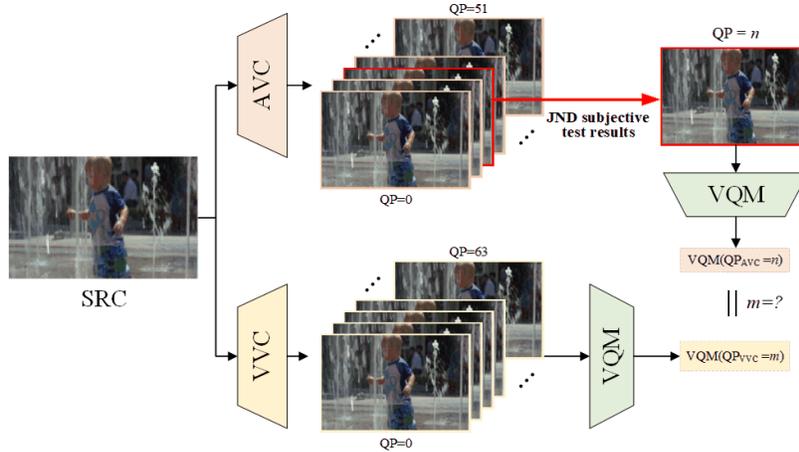


Figure 4: Mapping of the corresponding QP values from AVC to VVC.

2 Expanding JND Datasets to other Codecs

Carrying out a subjective test is both time-consuming and expensive. This is particularly true for JND subjective tests, as they involve the identification of JNDs within a large set of quality levels [13, 14]. VideoSet [12] is the most extensive JND dataset conducted on the AVC codec to determine the JND values for the QP proxy. The dataset comprises 220 source video sequences with a duration of 5 seconds, featuring frame rates of either 24 fps or 30 fps. These sequences were encoded at different resolutions using the constant quantization parameter (CQP) rate control mode of the

AVC, with QPs ranging from 0 to 51. The subjective evaluation of JND of individuals was conducted across multiple universities.

VideoSet currently supports only AVC and to address the need for JND datasets for more advanced codecs like HEVC and VVC, we aim to explore alternatives that do not require an extensive subjective testing process. Our objective is to extend the applicability of JND datasets to other codecs based on VideoSet, originally designed for AVC, while minimizing the need for additional subjective testing efforts. In Figure 4, we show the process of mapping the QP from subjective test results of AVC to VVC, all without the need for additional subjective tests. This mapping leverages Video Quality Metrics (VQMs) as proxies, operating under the assumption that widely adopted metrics like VMAF provide consistent scores for the same perceptual quality across different codecs for a given video. In the following section, we substantiate this assumption through a comprehensive analysis of existing cross-codec subjective datasets.

2.1 Codec-Agnostic Video Quality Metric (VQM)

We examine the per-content codec-agnostic features of various Video Quality Metrics (VQMs) by leveraging the subjective test results from the Waterloo IVC 4K dataset [17]. The Waterloo IVC 4K dataset includes subjective tests conducted on various video codecs, such as AVC, HEVC, VP9, and AV1. We proceed to compute widely used VQMs, including VMAF, PSNR, SSIM, and MS-SSIM, for all the encoded videos. Taking the mean opinion score (MOS) as the ground truth of perceptual quality, *for a VQM to demonstrate codec independence, it should yield the same VQM score for the same MOS across different codecs.*

To quantitatively evaluate the codec-independent performance of various VQMs, we employ a MOS-VQM regression analysis for each VQM. For a given video encoded with codec $c \in C$ at a VQM set of $\mathbf{Q}_c = [q_c^1, q_c^2, \dots, q_c^n]$, we represent the MOS of its encoded versions as $\mathbf{M}_c = [m_c^1, m_c^2, \dots, m_c^n]$.

The regression analysis aims to predict MOS values based on VQM scores for videos encoded using various codecs. Consequently, our goal is to identify the optimal VQM and its associated regression model that minimizes the mean absolute error (MAE) across all codecs:

$$\text{MAE} = \frac{1}{N} \sum_{c \in C} \sum_{i=1}^n |f(q_c^i) - m_c^i|, \quad (2)$$

where f is obtained from:

$$\min_f \|f(\mathbf{Q}_{c_1}) - \mathbf{M}_{c_1}\|_2^2. \quad (3)$$

This states that for the best fit of the model f , MAE between the predicted MOS values $f(\mathbf{Q}_c)$ and the actual MOS values \mathbf{M}_c over all codecs in set C should be minimal.

Given that subjective tests within the VideoSet dataset were performed on AVC only, our analysis selects AVC as c_1 from the codec set $C = \{\text{AVC}, \text{HEVC}, \text{VP9}, \text{AV1}\}$

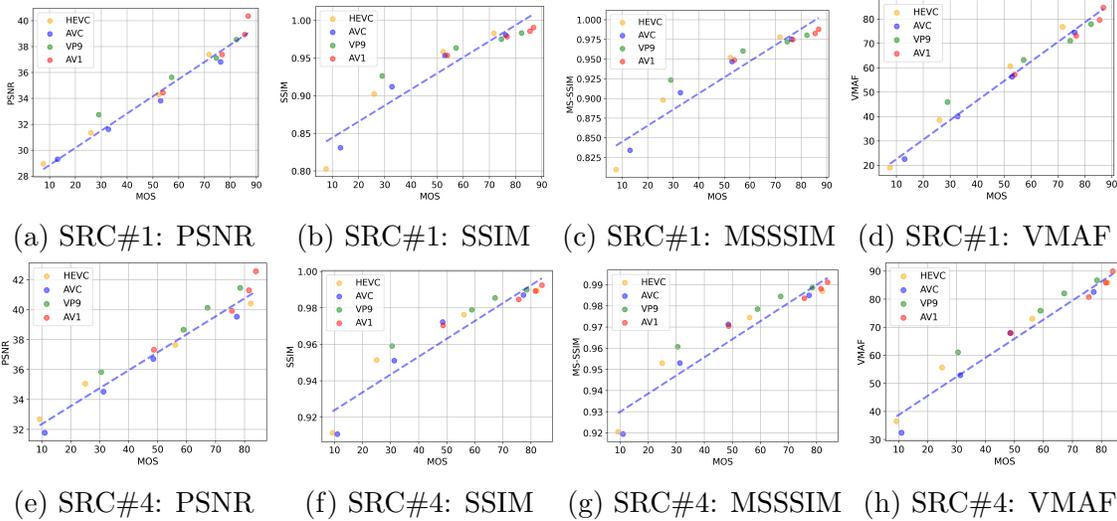


Figure 5: Relationships between MOS and different VQMs for SRC#1,4 in Waterloo IVC 4K datasets for different codecs

provided by the Waterloo IVC 4K dataset. For each content in the Waterloo IVC 4K dataset, we fit a regression model as per Equation (3), using the MOS scores associated with AVC encoded videos and their corresponding VQM scores. Figure 5 presents the plots of VQM versus MOS for two different content from the Waterloo IVC 4K dataset, using PSNR, SSIM, MS-SSIM, and VMAF as VQM. The fitted lines are derived from the VQM-MOS data of versions encoded with AVC using a linear regression model.

Table 1 summarizes the MAE values for the selected VQMs using the linear regression model. MAE measures the absolute difference between the predicted MOS values on the fitted line with AVC MOS-VQM points and the actual human-rated MOS scores. Lower MAE values indicate the codec-independence of the metric. The results indicate that, using linear regression, VMAF achieves the lowest MAE, at 4.51%, across all codecs, demonstrating its codec independence. This denotes that VMAF consistently estimates the MOS with a marginal error of 4.51% regardless of the codec used.

2.2 Generate Cross-Codec JND Datasets

We encoded the 220 source video of VideoSet with HEVC and VVC using the following configurations:

- HEVC: The x265 HEVC video encoder version 3.4, integrated with FFMPEG (libx265), was employed using its default settings (medium preset).
- VVC: The VVenC VVC encoder version 1.9.1 was operated using the ‘faster’ preset and default settings.

VideoSet is based on videos that are encoded with AVC with QP ranging from 0 to 51. For a given content m in VideoSet, assuming that there are N reliable subjects’

Table 1: MAE between MOS of different codecs and the curve fitted on the AVC codec on Waterloo dataset using linear regression model.

Codec	PSNR	SSIM	MS-SSIM	VMAF
AVC	2.86	6.28	5.73	3.62
HEVC	5.99	7.83	7.38	5.41
VP9	10.60	8.50	7.72	5.86
AV1	12.98	11.52	10.20	6.60
Average	8.11	8.53	7.76	4.51

JND annotations, the JND of a subject “n” is denoted by $j_n^{AVC,m}$. JND of N subjects can be denoted by $J^{AVC,m}$ as

$$J^{AVC,m} = [j_1^{AVC,m}, j_2^{AVC,m}, \dots, j_N^{AVC,m}] \quad (4)$$

For the same content m for another codec, taking VVC as example, the JND annotations of subject “n” can be computed by Equation (5).

$$j_n^{VVC,m} = \arg \min_{i \in \{1,2,\dots,63\}} |\text{VQM}(\text{QP}_{AVC}(j_n^{AVC,m})) - \text{VQM}(\text{QP}_{VVC}(i))|, \quad (5)$$

where $\text{QP}_c(x)$ is the video encoded with QP of x with codec c . This equation is based on the codec agnostic features of different VQMs which has been validated in Section 2.1.

3 Bitrate as a Function of Satisfied User Ratio (SUR)

The main objective of the study is to determine the additional bitrate expense required to improve user satisfaction when using various codecs. Accordingly, our initial step involves plotting the average bitrate for videos encoded at QPs corresponding to the $\text{SUR}(p\%)$ pertinent to different JND levels as a function of the user satisfaction ratio, *i.e.*, p . Figure 6a shows the average bitrates required to satisfy a certain user ratio using the AVC codec, as gathered directly from VideoSet. It has been observed that achieving higher user satisfaction—whereby a greater percentage of users cannot detect any distortion between the reference video and the compressed version—necessitates a greater bitrate. However, the required increase in bitrate to maintain user satisfaction exhibits an exponential trend. For example, to ensure that 75% of users cannot discern any difference at the 1st JND level, videos need to be encoded at an approximate average of 5 Mbps. If p is increased to 90%, the approximate average required bitrate escalates to 10 Mbps, which represents a significant increase. Figure 6b illustrates the increase in bitrate compared to the preceding satisfaction level for each given satisfaction ratio, p . It has also been noted that for the 1st JND, the increase in bitrate corresponding to an increase in the satisfaction ratio p is greater than that for the 2nd JND, and the increases for the 2nd JND are in turn higher than those for the 3rd JND.

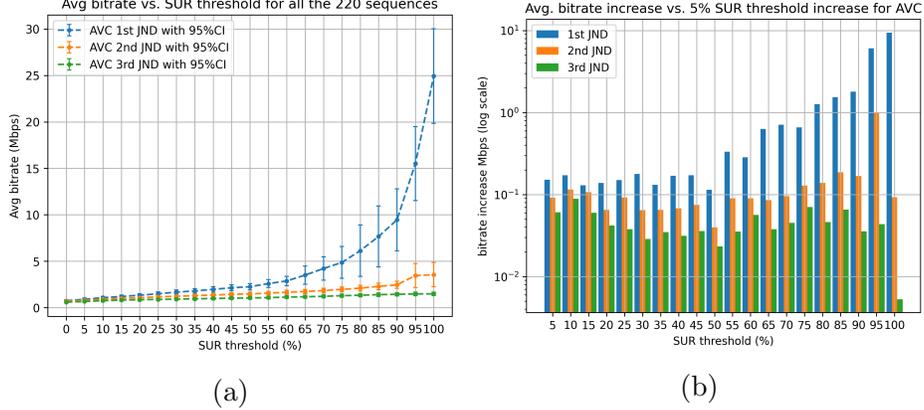


Figure 6: Relationships between bitrate and SUR threshold for VideoSet AVC 1080p on 1st, 2nd and 3rd JND. (a) Average bitrate for different SUR *thr.* (b) Bitrate increase for every 5% SUR *thr.* increase

3.1 Cross-Codec Comparison on SUR

After generating the new JND datasets for HEVC and VVC as described in Section 2.2, we conducted similar analyses to compare them with AVC. We normalized the average bitrate of HEVC and VVC at SUR($p\%$) by dividing them with the corresponding bitrate of AVC as shown in Equation (6).

$$Bitrate_{normalized}(SUR_{p\%}^c) = Bitrate(SUR_{p\%}^c) / Bitrate(SUR_{p\%}^{AVC}) \quad (6)$$

The results are shown in Figure 7a. It can be observed that, akin to AVC, the average bitrate required to satisfy a given percentage of users ($p\%$) increases with the $p\%$ value. However, the bitrate demand is comparatively lower for HEVC and VVC, as the relative bitrate is less than 1. An increase in relative bitrate can also be noted, suggesting that the efficiency advantages of HEVC and VVC over AVC are more pronounced at lower $p\%$ values, with this advantage diminishing as $p\%$ rises. The enhanced coding efficiency of HEVC and VVC in comparison to AVC accounts for the notable gains in bitrate. Figure 7b demonstrates the quantified relationship between VMAF scores and bitrates for the video SRC#1 encoded using different codecs. Additionally, it outlines the associated SUR at a 75% satisfaction level for the 1st, 2nd, and 3rd perceptual JND thresholds. The data clearly indicates that at identical JND levels, HEVC and VVC achieve the designated quality with lower bitrate requirements than AVC, confirming the superior performance of the former codecs.

3.2 Impact of Video Quality Metric

In Section 2.1, we established that VMAF exhibits the highest level of codec-agnostic properties compared to other VQMs. To assess the robustness of our JND mapping methodology across various codecs, we employ multiple VQMs to determine the JND points. This multi-metric approach enables us to understand the extent to which a VQM affects the mapping. Figure 8 shows the relative bitrate for the HEVC

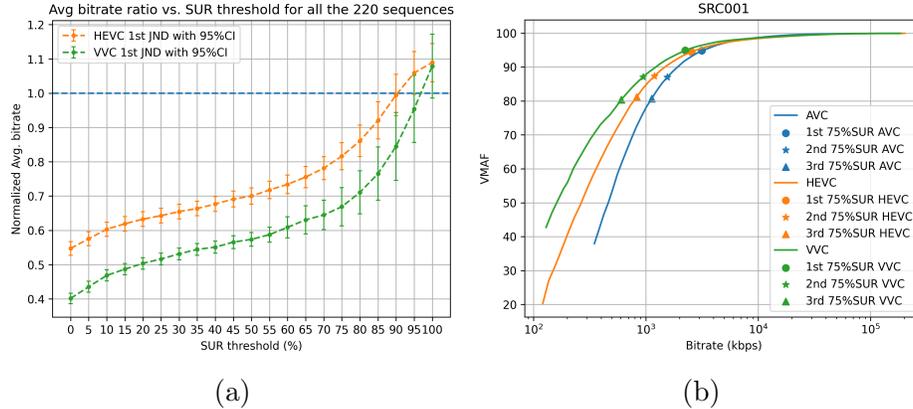


Figure 7: (a) The normalized average bitrate of HEVC and VVC at various SUR(p%)s. (b) VMAF vs. bitrate curves for video SRC# 1 encoded with AVC, HEVC, and VVC.

codec’s JND points using different VQMs for mapping. It has been observed that, aside from VMAF, other VQMs yield similar results for the 1st JND, however, with a significant margin of error compared to VMAF. However, for the 2nd and 3rd JNDs, the similarity between the results diminishes. This highlights the importance of choosing the appropriate VQM for accurate mapping.

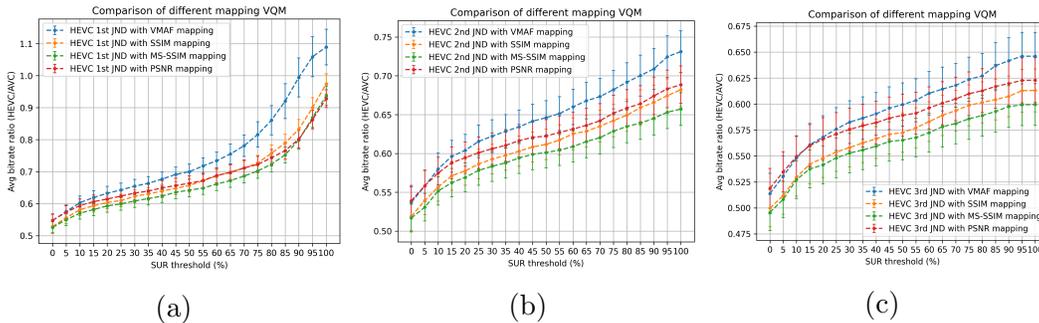


Figure 8: Comparative mapping results to HEVC videos using different VQMs.

4 Conclusion

In this study, we conducted an evaluation of the codec-agnostic characteristics of various video quality metrics. Our findings indicate that VMAF exhibits superior robustness across different codecs, meaning that an equivalent VMAF score for different codecs correlates with a comparable Mean Opinion Score (MOS) for those codecs. Utilizing VMAF, we mapped Just Noticeable Difference (JND) points from the AVC codec within the VideoSet dataset to the HEVC and VVC codecs, based on the premise that equivalent VMAF scores for two codecs correlate to the same perceptual video quality. Subsequently, we assessed the bitrate investment necessary to achieve an increased number of satisfied users across different codecs. Our observations reveal an exponential trend, suggesting that satisfying a larger proportion of users who cannot detect the quality differences between a reference video and its encoded counterpart necessitates a step increase in bitrate.

References

- [1] Abdelhak Bentaleb, Bayan Taani, Ali C. Begen, Christian Timmerer, and Roger Zimmermann, “A Survey on Bitrate Adaptation Schemes for Streaming Media Over HTTP,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 562–585, 2019.
- [2] Hadi Amirpour, Christian Timmerer, and Mohammad Ghanbari, “PSTR: Per-Title Encoding Using Spatio-Temporal Resolutions,” in *IEEE ICME*, July 2021, pp. 1–6.
- [3] Apple, “HTTP Live Streaming (HLS) Authoring Specification for Apple Devices | Apple Developer Documentation,” 2015.
- [4] Jan De Cock, Zhi Li, Megha Manohara, and Anne Aaron, “Complexity-Based Consistent-Quality Encoding in the Cloud,” in *IEEE ICIP*, Sept. 2016, pp. 1484–1488.
- [5] Hadi Amirpour, Mohammad Ghanbari, and Christian Timmerer, “DeepStream: Video Streaming Enhancements using Compressed Deep Neural Networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2022.
- [6] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara, “Toward A Practical Perceptual Video Quality Metric,” *The Netflix Tech Blog*, vol. 6, pp. 2, 2016.
- [7] Angeliki V. Katsenou, Fan Zhang, Kyle Swanson, Mariana Afonso, Joel Sole, and David R. Bull, “VMAF-Based Bitrate Ladder Estimation for Adaptive Streaming,” in *PCS*, June 2021, pp. 1–5.
- [8] Yuriy A. Reznik, Karl O. Lillevold, Abhijith Jagannath, Justin Greer, and Jon Corley, “Optimal Design of Encoding Profiles for ABR Streaming,” in *PV Workshop*, June 2018, pp. 43–47.
- [9] Farzad Tashtarian, Abdelhak Bentaleb, Hadi Amirpour, Babak Taraghi, Christian Timmerer, Hermann Hellwagner, and Roger Zimmermann, “LALISA: Adaptive Bitrate Ladder Optimization in HTTP-based Adaptive Live Streaming,” in *IEEE/IFIP NOMS*. May 2023, pp. 1–9, IEEE.
- [10] Vignesh V Menon, Hadi Amirpour, Mohammad Ghanbari, and Christian Timmerer, “Perceptually-Aware Per-Title Encoding for Adaptive Video Streaming,” in *IEEE ICME*, July 2022, pp. 1–6.
- [11] Madhukar Bhat, Jean-Marc Thiesse, and Patrick Le Callet, “Combining Video Quality Metrics To Select Perceptually Accurate Resolution In A Wide Quality Range: A Case Study,” in *IEEE ICIP*, Sept. 2021, pp. 2164–2168.
- [12] Haiqiang Wang, Ioannis Katsavounidis, Jiantong Zhou, Jeonghoon Park, Shawmin Lei, Xin Zhou, Man-On Pun, Xin Jin, Ronggang Wang, Xu Wang, Yun Zhang, Jiwu Huang, Sam Kwong, and C. C. Jay Kuo, “VideoSet: A Large-scale Compressed Video Quality Dataset based on JND Measurement,” *Journal of Visual Communication and Image Representation*, vol. 46, pp. 292–302, July 2017.
- [13] Jingwen Zhu, Suiyi Ling, Yoann Baveye, and Patrick Le Callet, “A Framework to Map VMAF with the Probability of Just Noticeable Difference between Video Encoding Recipes,” in *IEEE IVMSIP*. IEEE, 2022, pp. 1–5.
- [14] Mohsen Jenadeleh, Raouf Hamzaoui, Ulf-Dietrich Reips, and Dietmar Saupe, “Crowdsourced estimation of collective just noticeable difference for compressed video with flicker test and quest+,” 2023.
- [15] Jingwen Zhu, Patrick Le Callet, Anne-Flore Perrin, Sriram Sethuraman, and Kumar Rahul, “On the benefit of parameter-driven approaches for the modeling and the prediction of satisfied user ratio for compressed video,” in *IEEE ICIP*. IEEE, 2022, pp. 4213–4217.
- [16] Xinfeng Zhang, Chao Yang, Haiqiang Wang, Wei Xu, and C.-C. Jay Kuo, “Satisfied-User-Ratio Modeling for Compressed Video,” *IEEE Transactions on Image Processing*, vol. 29, pp. 3777–3789, 2020, Conference Name: IEEE Transactions on Image Processing.
- [17] Zhuoran Li, Zhengfang Duanmu, Wentao Liu, and Zhou Wang, “AVC, HEVC, VP9, AVS2 or AV1? — A Comparative Study of State-of-the-Art Video Encoders on 4K Videos,” in *Image Analysis and Recognition*, vol. 11662, pp. 162–173. Springer International Publishing, Cham, 2019, Series Title: Lecture Notes in Computer Science.