applied sciences

MDPI

Article

Assessment of the Quality of Video Sequences Performed by Viewers at Home and in the Laboratory

Janusz Klink [1,*], Stefan Brachmański [2] and Michał Łuczyński [2]

[1] Department of Telecommunications and Teleinformatics, Faculty of Information and Telecommunication Technology, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
[2] Photonics and Microsystems Department of Acoustics, Multimedia and Signal Processing, Faculty of Electronic, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
* Correspondence: janusz.klink@pwr.edu.pl

Featured Application: The results of the research may be helpful in the setup of video quality assessment procedures in order to achieve results as close as possible to the quality experienced by the end users of the video streaming services.

Abstract: The paper presents the results of subjective and objective quality assessments of H.264-, H.265-, and VP9-encoded video. Most of the literature is devoted to subjective quality assessment in well-defined laboratory circumstances. However, the end users usually watch the films in their home environments, which may be different from the conditions recommended for laboratory measurements. This may cause significant differences in the quality assessment scores. Thus, the aim of the research is to show the impact of environmental conditions on the video quality perceived by the user. The subjective assessment was made in two different environments: in the laboratory and in users' homes, where people often watch movies on their laptops. The video signal was assessed by young viewers who were not experts in the field of quality assessment. The tests were performed taking into account different image resolutions and different bit rates. The research showed strong correlations between the obtained results and the coding bit rates used, and revealed a significant difference between the quality scores obtained in the laboratory and at home. As a conclusion, it must be underlined that the laboratory tests are necessary for comparative purposes, while the assessment of the video quality experienced by end users should be performed under circumstances that are as close as possible to the user's home environment.

Keywords: video quality; objective quality assessment methods; subjective quality assessment methods; user experience; user perception; QoE; video codecs; H.264 (AVC); H.265 (HEVC); VP9

Citation: Klink, J.; Brachmański, S.; Łuczyński, M. Assessment of the Quality of Video Sequences Performed by Viewers at Home and in the Laboratory. *Appl. Sci.* **2023**, *13*, 5025. https://doi.org/10.3390/app13085025

Academic Editor: Yutaka Ishibashi

Received: 27 February 2023
Revised: 4 April 2023
Accepted: 14 April 2023
Published: 17 April 2023

1. Introduction

For many years, in everyday life, television played the role of the most important medium. Significant changes are currently being observed, especially among the young generation. Today's youth are increasingly willing to watch TV broadcasts, including movies, via the Internet using mobile devices or laptops. An important issue is the quality of the video delivered to the users. There are two general approaches to quality assessment, namely, subjective and objective. The recommendations of the International Telecommunication Union (ITU) [1] specify, in detail, the conditions for performing measurements related to the subjective assessment of video quality. In general, the assessment should be conducted under laboratory conditions, possibly simulating home conditions. The universal measurement room should be able to meet the requirements of an idealized room as well as a domestic room. Recommendation BT.500 [1] defines the general viewing conditions for subjective assessments in a laboratory and in the home environment. However, it should be taken into account that the general viewing conditions for the home environment do not guarantee that they will suit the specific home conditions of each user. An Internet user usually watches a video transmission under conditions that do not always meet the requirements of the ITU BT.500 recommendation. Consequently, the evaluation of video quality performed under real home conditions and under home conditions emulated in the laboratory will not necessarily be the same. Moreover, conducting the research in real users' locations allows us to spread the test environment across a much wider population of service customers. In general, the viewing conditions can significantly impact the results obtained. The purpose of the evaluation and the audience to whom the evaluation is devoted may determine the acceptable circumstances of the test. The next issue of such a subjective evaluation is its high cost, because many factors must be included in the experimental design and many subjects (human testers) must be involved. Therefore, many studies try to find a replacement for these methods by modeling, simulating the real world in an artificial environment, or using objective approaches to quality assessment [2]. In the next step, different approaches to quality modeling may be applied, taking into account the different points of view of various stakeholders in the media streaming process [3]. However, the results obtained from objective methods may not always correlate with subjective users' scores, especially when the circumstances of subjective quality assessment are changing. The authors anticipate that the video quality assessment performed in the artificial environment (even if the conditions emulate an 'average' home) may give different results from the scores obtained under real home conditions. Examining this may provide an answer as to whether the assessment of video quality experienced by users may (or may not) always be replaced by laboratory tests. The next issue occurs when we talk about comparisons between the subjective and objective results of the quality assessment. This is not a trivial task, taking into account that there are plenty of objective methods and quality metrics. There is a long list of literature that describes their characteristics. Some studies discuss their strengths and weaknesses, as well as their usability to predict video quality assessed by end users, especially when talking about metrics such as mean squared error (MSE) and peak signal-to-noise ratio (PSNR) [4–6]. Others show that the correspondence between objective and subjective scores also depends on the video content and show that some metrics, such as the structural similarity (SSIM) index, present characteristics similar to the human visual system (HVS) and their results are closer to users' subjective scores [7,8]. Finally, there are papers that give a comprehensive view of the different factors that influence the degradation of the video content delivered to the user, present a broad review of objective video quality assessment methods, their classification, and performance comparison [9], and give a survey of the evolution of these methods, analyzing their characteristics, advantages, and drawbacks [10,11]. Most of them are good enough for comparison and benchmarking purposes [12–15], but some give results that present stronger correlations with quality of experience (QoE) scores, given by users during subjective quality assessment, than others. Mapping the quality of service (QoS) onto QoE allows us to build proper QoE models. However, finding general relationships between QoS and QoE is not an easy task. Sometimes, the content of the video may influence the perceptual-based quality assessment in specific circumstances [16–19]. This is why big content providers and streaming platforms, which use Dynamic Adaptive Streaming over HTTP (DASH) mechanisms to provide their content via the Internet, use different coding bit rate ladders according to the video content provided [20,21]. Furthermore, the bit rate coding ladder for specific video content may depend on the video codec [22]. The authors chose three objective quality metrics, namely, the PSNR [23], SSIM [24], and the video multimethod assessment fusion (VMAF) [25], from the long list of metrics proposed in the literature. The PSNR metric is often used because it has a clear physical meaning and is simple to calculate. It presents good results when assessing the influence of some degradation factors on the quality of specific video footage, e.g., before and after compression. However, it may not always be sufficiently correlated with subjective quality assessment scores. PSNR is memoryless, which means that it is calculated pixel by pixel, independently, for each pair of corresponding frames of the two compared videos and assumes that the video quality is independent of the spatial and temporal relationships between the samples of the source footage. Reordering the pixels in the reference and examined videos, in the same way, does not change the PSNR values, although the subjective quality may change. Moreover, it can be found in the literature that video signals are highly structured and the ordering of pixels carries important perceptual structural information about the contents of the visual scene [4]. This led to also taking into account other video quality metrics, such as SSIM and VMAF, which take into account the fact that natural image signals are highly structured and may possibly better correlate with subjective quality assessment scores [11,24–26]. When there is no possibility to access the original footage, then no-reference (NR) image or video quality assessment methods can be used to evaluate the quality of the material delivered to the end user. The original footage may be distorted at any stage of the media delivery chain, that is, during acquisition, processing, compression, transmission, decoding, or presentation at the receiver's site. Therefore, it is important to use quality assessment methods that are based on a good representation of different types of distortions and may use them for a proper evaluation. Early NR quality assessment methods usually took into account specific distortion types, such as blur [27], blocking [28], and ringing artifacts [29]. In real situations, the distortion types are usually not known in advance; thus, recently, more attention has been paid to general-purpose NR methods. These metrics attempt to learn the knowledge when evaluating the quality of images and characterize the general rules of image distortions. On the basis of this knowledge, image quality prediction models can be established and adapted to unknown distortions [30]. There are many approaches based on deep convolutional neural networks (DCNN) NR image quality assessment (IQA) [31–33]. They emphasize a good distortion representation, which is crucial for the performance of NR-IQA or blind image quality assessment (BIQA). In [34], the relationship between different distortion levels and their types is analyzed. The authors proposed a new approach, named 'GraphIQA', which presents a distortion graph representation-based deep learning BIQA. General-purpose BIQA models suffer from catastrophic forgetting, which refers to the tendency of a neural network to 'forget'. A solution to this problem may be the lifelong blind image quality assessment (LIQA) approach, which not only learns new distortions, but can also mitigate the catastrophic forgetting of identified distortions [35]. The main purpose of our work was to assess the influence of the environment on the video quality experienced by the user and to find correlations with the results of the objective quality assessment. The objective evaluation was based on the full reference (FR) method, where not only the distorted video, but also the reference footage was available.

The goals of the research were to:

1. Conduct a comparative analysis of the video quality assessment results obtained under laboratory and real home (not lab-emulated) conditions;
2. Find correlations between objective results and subjective assessment scores, taking into account the influence of the test environment.

The results of the research should answer the question of whether laboratory tests can replace the video quality assessment conducted in users' homes and reduce testing costs. Furthermore, the research should show which type of subjective quality assessment is more closely correlated with objective quality assessment methods and which metric is worth using.

The video quality assessment was made taking into account:

1. H.264, H.265, and VP9 encodings [36–38];
2. The bit rate (from 300 kbps to 6000 kbps);
3. Resolutions (640 × 360—ninth high definition (nHD), 858 × 480—standard definition (SD), 1280 × 720—high definition (HD), and 1920 × 1080—full high definition (Full HD)).

The paper is organized as follows. After the introduction, Section 2 describes the video test sample preparation procedure and the methods used in the research. In the next section, the results of the subjective and objective quality assessment are presented and discussed. At the end, the results are summarized and the conclusions drawn.

2. Materials and Methods

The first step of the research consisted of using the subjective video quality assessment. From many different video quality assessment methods [1,39–42], the comparative method Double Stimulus Impairment Scale Method (DSSM) was used in the study. The DSSM is recommended by the International Telecommunication Union (ITU), and the measurement technique is described in the BT.500 recommendation [1]. The evaluation consists of comparing the reference video sequence (reference signal) with the evaluated sequence. The reference signal was presented first and assessed second. The task of the observer (viewer) was to assess the degree of deterioration of the second signal in relation to the first signal. The rating was given on a five-point mean opinion score (MOS) scale, where 5 means invisible quality deterioration, 4—noticeable but not annoying, 3—slightly annoying, 2—annoying, and 1—very annoying [39]. The video sequences were presented to the observers in single pairs (pattern-evaluated sequence). Each pair was assessed separately. The reference and evaluated video sequences were separated by a gray screen presented to the observers for about 2 s.

Measurements were made for two cases:

1. Evaluation in the laboratory;
2. Ratings at the viewer's home.

The evaluation of video quality for condition 1, that is, in the laboratory, was carried out in a room adapted for the evaluation of video signals, equipped with a 60-inch TV screen. The laboratory room met the requirements of the recommendations of the International Telecommunication Union [1,39,43,44]. Its additional advantage was the fact that all participants in the research knew about it, so it did not affect the distraction of the students related to the adaptation to the location of the research. In turn, the video quality assessment for Case 2 was made in home conditions, i.e., not ideal, but still ensuring the quality assessment by the consumer. All participants in the measurements evaluated the video sequence on high-definition television (HDTV) monitors with a resolution of 1920×1080 [45]. The standard test material was a 20 s video sequence (without sound) with a resolution of 1920×1080 pixels in AVI format. The length of the video footage was twice as long as the (minimum) value proposed in [1]. This decision does not negatively affect the results of the subjective assessment, but, in the case of objective evaluation, allows one to calculate quality metrics based on a larger dataset, which, in the case of films with varied dynamics and content of the presented scenes, may positively influence the proper calculation of objective quality metrics, which will be more representative and more correlated with the subjective assessment. However, there are studies that take into account longer video samples. This case was described in [46], where the authors considered 180 s samples for the evaluation of QoE in adaptive video streaming over wireless networks. Longer samples allow for a better evaluation of the quality perceived by users, especially when transmission disturbances may occur irregularly and at relatively longer intervals. The test footage included horse racing start scenes (see Figure 1) [47].

The original sequence was encoded in H.264, H.265, and VP09 with different resolutions and different bit rates. Four resolutions were taken into account in the research: 640×360 (360p), 858×480 (480p), 1280×720 (720p), and 1920×1080 (1080p). For the coding techniques and a specific resolution, various transmission conditions were simulated with 18 bit rates: 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, and 6000 kbps. The test material was presented to viewers divided into encoding technique and resolution. The test videos with different transmission conditions (bit rate) were randomly presented to the viewers. Each group of viewers evaluated the video signal subjected to one encoding technique for all resolutions and bit rates. In both cases, the team of observers consisted of second-year electronics students at the Faculty of Electronics, Photonics, and Microsystems of the Wrocław University of Science and Technology, aged 20–21, with normal visual acuity and correct color discrimination. As recommended by the International Telecommunication Union BT. 500 [1], the minimum number of observers should be 15. In the presented studies, three groups were created for home measurements and three groups for laboratory measurements. Each group evaluated a different type of coding. The number of individual test groups examining individual codecs was as follows:

1. H.264—25 people under home conditions and 45 people under laboratory conditions;
2. H.265—35 people under home conditions and 35 people under laboratory conditions;
3. VP09—30 people under home conditions and 40 people under laboratory conditions.

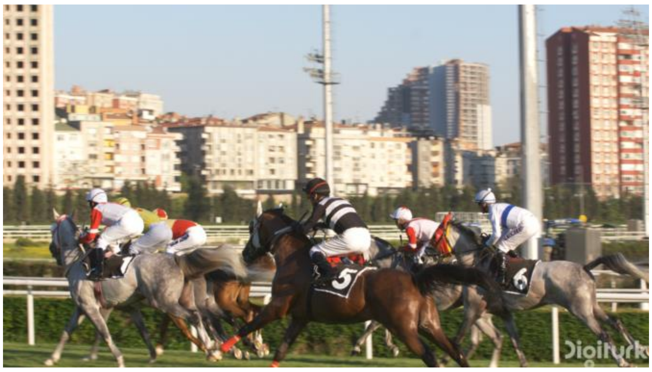

Figure 1. An example frame from the original video.

The different size of the groups was, among other reasons, the result of a different number of individuals willing to participate in a given measurement session, the effect of a statistical analysis of observer ratings, and the elimination of those observers who were characterized by low participation (regardless of the bit rate or resolution, they gave the same quality rating). Before beginning the measurements, the participants were familiarized with the assessment method and had one training session. During the training, the observers became acquainted with the technique of presenting the test material and how to assess the changes in video quality. After the training, the actual measurements began. After viewing the original and encoded sequences, each study participant recorded their assessment of quality deterioration in a special form. In the second part of the research, the authors performed video quality assessment using the objective double stimulus method, which relies on a comparison of the encoded video samples with the reference original video (see Figure 2).

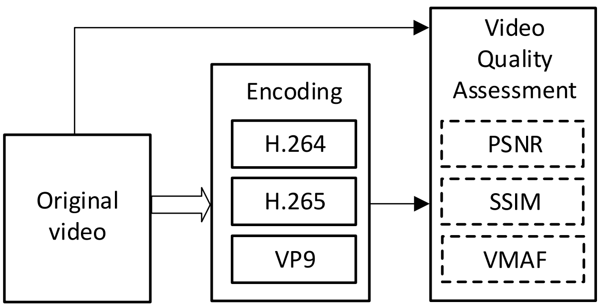

Figure 2. Video quality assessment using double stimulus method.

The original video footage was encoded using the FFmpeg [48] tool with implemented H.264, H.265, and VP9 video codecs. Four spatial resolutions and coding bit rates in the range from 300 to 6000 kbps, mentioned above, were taken into account. There were 216 video samples (3 codecs × 4 spatial resolutions × 18 coding bit rates) prepared in total.

Each set of video samples for specific spatial resolution should be compared with source video footage of the same resolution. This way, the quality of each set of videos is objectively assessed independently of the other sets of videos. When it comes to subjective quality assessment, each set of videos should be presented on a display with proper resolution that should be fitted to the resolution of the assessed video. This may be difficult to achieve when the quality assessment is performed by many different users in their home environments, where a specific display resolution may be used by default. Thus, the authors assumed that the objective assessment should be conducted using one display resolution. The most popular spatial resolution of the displays used by end users was 1920 × 1080 pixels (FHD). Therefore, the sample preparation process was a bit more complicated than just encoding. It also included upsizing all of the videos of smaller spatial resolution, i.e., 640 × 360, 858 × 480 and 1280 × 720, to FHD (Figure 3). This way, the authors wanted to achieve the same effect observed on the end-user equipment, which usually resizes smaller resolution videos to the maximum display size, with FHD resolution set by default.

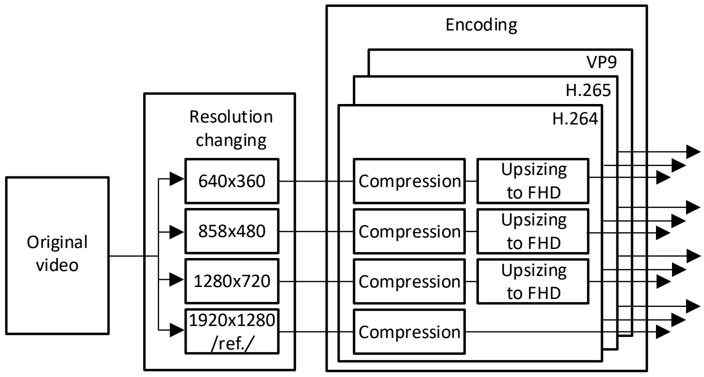

Figure 3. Video sample preparation procedure.

After this preparation, the tested video samples were objectively assessed, by comparison with the reference video (denoted in Figure 3 as '1920 × 1280/ref./'), using three metrics, i.e., PSNR, SSIM, and VMAF. Finally, these results could be compared with the subjective user scores. A detailed description of the methodology in the form of a flow diagram of the work is presented in Figure 4.

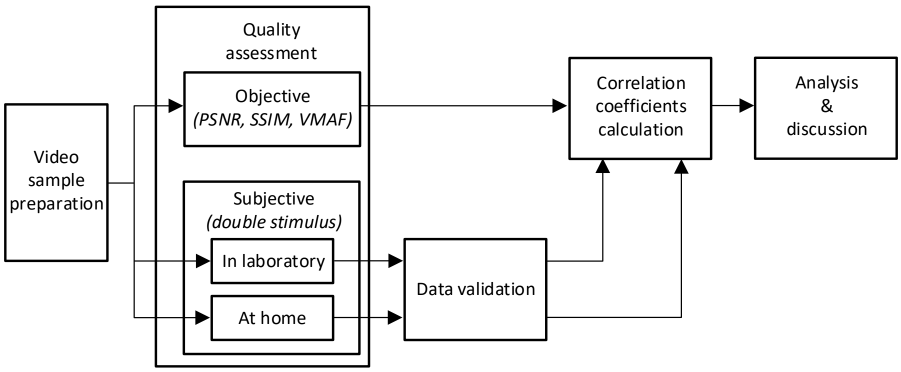

Figure 4. Flow diagram of the work.

3. Results and Discussion

The results of the subjective assessment of the quality of the video were entered into a spreadsheet and subjected to statistical analysis according to the procedure described in the ITU-R BT.500 recommendation [1]. In accordance with this recommendation, a 95% confidence interval was adopted. The mean value of the MOS score in the group of observers was calculated separately for each encoding technique, screen resolution, and bit rate.

3.1. Subjective Quality Assessment of Video Encoded Using H.264 Standard

The H.264 standard [36], also known as MPEG-4 Part 10 or AVC (advanced video coding), was introduced in 2003 as a result of cooperation between the ITU-T Q.6/SG16 Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG). The team formed in this way is known as the Joint Video Team (JVT). The H.264 standard uses differential compression, in which the current image is created based on one or more previous images, taking into account the differences that occurred between them at that time. In relation to earlier solutions, the H.264 standard uses a number of improvements, on the one hand, allowing a reduction in the bit rate with unchanged image quality, and on the other hand, significantly increasing the demand for computing power during encoding. The degradation of the quality of the video signal encoded in the H.264 standard was assessed in a group of 25 people at home and 45 people in a laboratory. The obtained results of the measurements made under home conditions are presented in Table 1 and graphically in Figure 5, with the laboratory conditions presented in Table 2 and Figure 6. In addition to the MOS mean value, the tables also include the standard deviation values (S), as well as the values of the confidence interval coefficient (δ) calculated according to the ITU BT.500 recommendation [1].

Table 1. Mean value of the video quality assessment (MOS) for H.264 codec, standard deviation (S), and confidence interval coefficient (δ) for four resolutions—measurements at home [49].

Bit Rate	640 × 360			858 × 480			1280 × 720			1920 × 1080		
(kbps)	MOS	S	δ	MOS	S	δ	MOS	S	δ	MOS	S	δ
300	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.05	0.23	0.10
400	1.11	0.32	0.15	1.21	0.42	0.19	1.26	0.45	0.20	1.26	0.45	0.20
500	1.50	0.51	0.24	1.42	0.69	0.31	1.58	0.69	0.31	1.42	0.51	0.23
600	1.74	0.73	0.33	1.72	0.57	0.27	1.84	0.60	0.27	1.68	0.48	0.21
700	1.94	0.73	0.34	2.11	0.58	0.27	2.11	0.57	0.26	1.89	0.32	0.14
800	2.33	0.49	0.22	2.26	0.65	0.29	2.39	0.70	0.32	2.21	0.54	0.24
900	2.42	0.51	0.23	2.47	0.62	0.30	2.61	0.70	0.32	2.58	0.51	0.23
1000	2.71	0.92	0.44	2.79	0.54	0.24	2.83	0.62	0.29	2.84	0.50	0.23
1500	2.89	0.88	0.39	3.00	0.49	0.22	3.16	0.76	0.34	3.37	0.50	0.22
2000	2.95	0.71	0.32	3.11	0.68	0.31	3.50	0.51	0.24	3.79	0.63	0.28
2500	3.00	0.67	0.30	3.24	0.75	0.36	3.67	0.59	0.27	4.06	0.73	0.34
3000	3.05	0.62	0.28	3.33	0.77	0.35	3.88	0.78	0.37	4.28	0.67	0.31
3500	3.17	0.62	0.29	3.42	0.77	0.35	4.06	0.68	0.33	4.41	0.51	0.24
4000	3.22	0.55	0.25	3.58	0.84	0.38	4.11	0.58	0.27	4.56	0.51	0.24
4500	3.33	0.59	0.27	3.68	0.89	0.40	4.19	0.66	0.32	4.67	0.49	0.22
5000	3.38	0.72	0.35	3.84	0.76	0.34	4.25	0.58	0.28	4.78	0.43	0.20
5500	3.44	0.62	0.28	3.95	0.71	0.32	4.32	0.58	0.26	4.83	0.38	0.18
6000	3.57	0.65	0.34	4.06	0.73	0.34	4.37	0.50	0.22	4.94	0.24	0.11

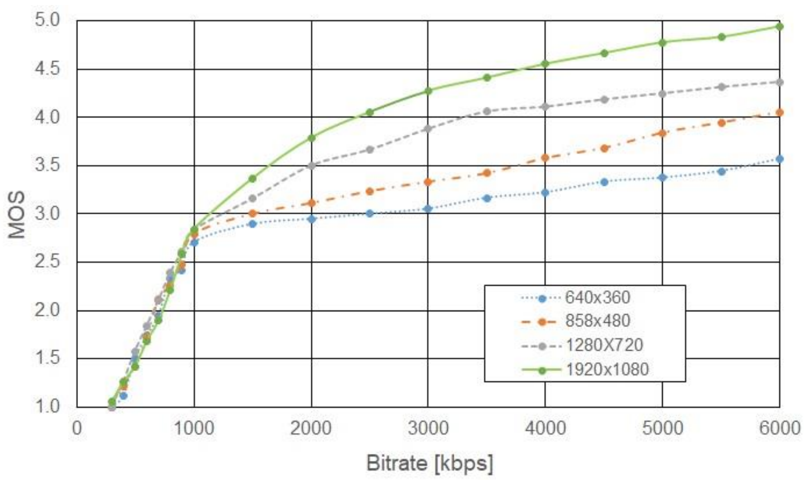

Figure 5. Results of the subjective quality assessment (MOS) for the H.264-encoded video as a function of bit rate for different spatial resolutions—measurements at home [49].

Table 2. Mean value of the video quality assessment (MOS) for H.264 codec, standard deviation (S), and confidence interval coefficient (δ) for four resolutions—measurements in laboratory.

Bit Rate	640 × 360			858 × 480			1280 × 720			1920 × 1080		
(kbps)	MOS	S	δ	MOS	S	δ	MOS	S	δ	MOS	S	δ
300	1.33	0.48	0.15	1.21	0.42	0.13	1.16	0.43	0.13	1.05	0.32	0.10
400	1.54	0.51	0.16	1.51	0.55	0.16	1.33	0.57	0.17	1.19	0.45	0.13
500	1.85	0.71	0.22	1.95	0.49	0.15	1.81	0.55	0.16	1.35	0.57	0.17
600	2.18	0.79	0.25	2.26	0.62	0.19	2.24	0.66	0.20	1.58	0.59	0.18
700	2.69	0.52	0.16	2.63	0.66	0.20	2.71	0.56	0.17	1.98	0.64	0.19
800	2.90	0.55	0.17	2.95	0.49	0.15	2.86	0.64	0.19	2.40	0.54	0.16
900	2.97	0.49	0.15	3.00	0.62	0.19	3.12	0.54	0.16	2.81	0.55	0.16
1000	3.08	0.62	0.20	3.16	0.69	0.21	3.26	0.54	0.16	3.16	0.57	0.17
1500	3.38	0.49	0.15	3.60	0.69	0.21	3.74	0.66	0.20	3.86	0.47	0.14
2000	3.49	0.64	0.20	3.81	0.70	0.21	4.07	0.70	0.21	4.26	0.49	0.15
2500	3.54	0.79	0.25	3.91	0.53	0.16	4.19	0.59	0.18	4.40	0.54	0.16
3000	3.59	0.64	0.20	4.05	0.58	0.17	4.28	0.55	0.16	4.49	0.51	0.15
3500	3.61	0.72	0.23	4.07	0.74	0.22	4.33	0.47	0.14	4.54	0.55	0.17
4000	3.68	0.66	0.21	4.14	0.74	0.22	4.40	0.49	0.15	4.63	0.49	0.15
4500	3.74	0.60	0.19	4.21	0.60	0.18	4.49	0.51	0.15	4.70	0.46	0.14
5000	3.79	0.77	0.24	4.26	0.62	0.19	4.56	0.50	0.15	4.79	0.41	0.12
5500	3.82	0.51	0.16	4.36	0.48	0.15	4.65	0.48	0.14	4.84	0.37	0.11
6000	3.85	0.49	0.15	4.40	0.49	0.15	4.72	0.45	0.14	4.91	0.29	0.09

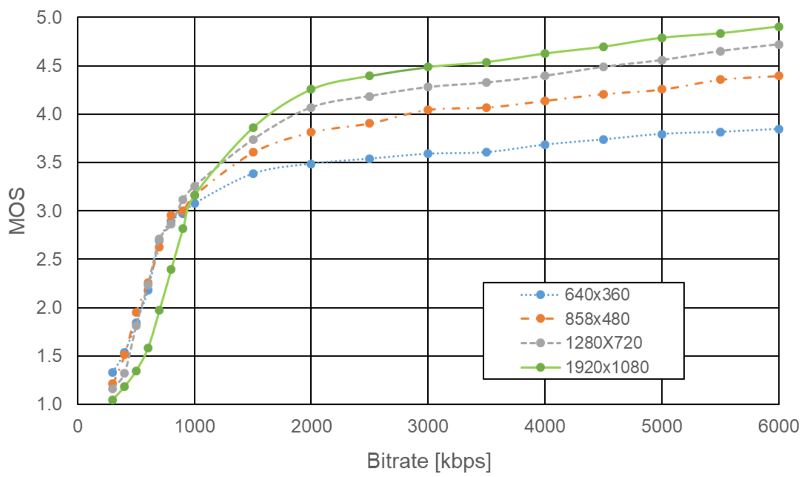

Figure 6. Results of the subjective quality assessment (MOS) for the H.264-encoded video as a function of bit rate for different spatial resolutions—measurements in the laboratory.

The statistical analysis of the results showed that up to a bit rate of 1000 kbps, the resolution does not affect the assessment of the video image quality made at home, while the laboratory measurements show a slightly lower assessment for the resolution of 1920 × 1080. Above this bit rate, the video quality depends on the resolution, and as expected, the video signal with a resolution of 1920 × 1080 is the highest rated. For this resolution, an MOS rating of at least 4 was achieved in home conditions at a bit rate starting from 2500 kbps and in laboratory conditions from approximately 1500 kbps. On the other hand, for a resolution of 1280 × 720, the MOS value of 4 was achieved at home at 3500 kbps and in laboratory conditions at 2000 kbps. Comparing the results of the MOS evaluation obtained in the laboratory and home conditions, it can be seen that viewers rated the video image presented under laboratory conditions more highly; only for the highest resolution at bit rates up to 600 kbps was the opposite MOS result found. Table 3 and Figure 7 show the difference of ΔMOS in the evaluation of video quality obtained for measurements made in the laboratory and home conditions according to Formula (1):

$$\Delta MOS = MOS_L - MOS_H, \quad (1)$$

where MOS_L is the evaluation obtained under laboratory conditions and MOS_H is the evaluation obtained under home conditions.

Table 3. Value of the difference in ΔMOS (for H.264 codec) between the results obtained under laboratory and home conditions.

Bit Rate	**ΔMOS**			
(kbps)	**640 × 360**	**858 × 480**	**1280 × 720**	**1920 × 1080**
300	0.33	0.21	0.16	0.00
400	0.43	0.30	0.06	−0.08
500	0.35	0.53	0.24	−0.07
600	0.44	0.53	0.40	−0.10
700	0.75	0.52	0.60	0.08
800	0.56	0.69	0.47	0.18
900	0.55	0.53	0.51	0.24
1000	0.37	0.37	0.42	0.32
1500	0.49	0.60	0.58	0.49
2000	0.54	0.70	0.57	0.47
2500	0.54	0.67	0.52	0.34
3000	0.54	0.71	0.40	0.21
3500	0.44	0.65	0.26	0.13
4000	0.46	0.56	0.28	0.07
4500	0.40	0.53	0.30	0.03
5000	0.42	0.41	0.31	0.01
5500	0.37	0.41	0.34	0.00
6000	0.27	0.34	0.35	−0.04

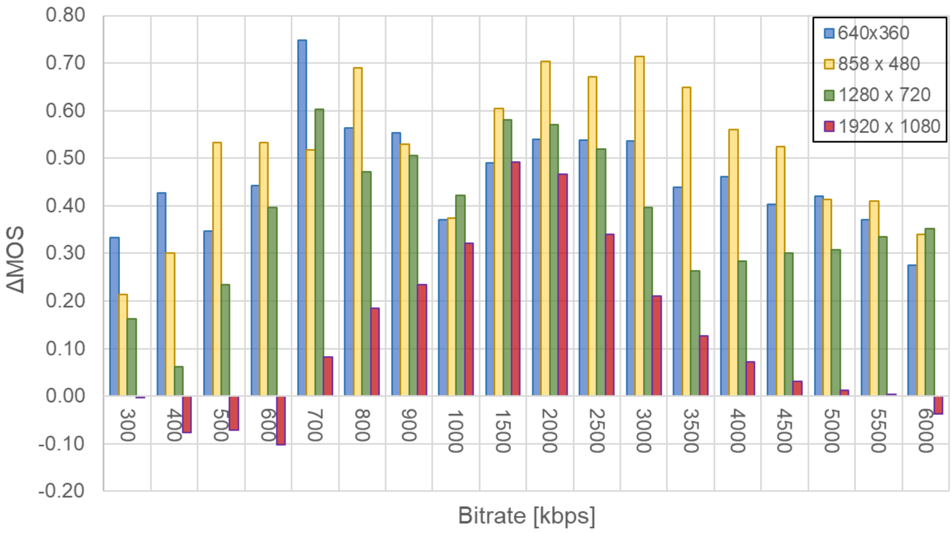

Figure 7. Difference between the results of the subjective evaluation of the H.264-encoded video (ΔMOS), conducted in the laboratory and at home, as a function of bit rate for different resolutions.

The statistical analysis of the results obtained using the *t*-test showed that at the confidence level of $\alpha = 0.05$, there is no basis to accept the hypothesis of the identity of the results obtained under laboratory and home conditions. The *t*-test values obtained using the Statistica tool for each resolution are as follows:

- 640 × 360; $t = 17.6 > t\ \alpha = 2.1$, at $\alpha = 0.05$;
- 858 × 480; $t = 14.9 > t\ \alpha = 2.1$, at $\alpha = 0.05$;
- 1280 × 720; $t = 10.7 > t\ \alpha = 2.1$, at $\alpha = 0.05$;
- 1920 × 1080; $t = 2.9 > t\ \alpha = 2.1$, with $\alpha = 0.05$.

It can be concluded that the differences between the MOS values obtained under the laboratory and home conditions are significant.

3.2. Subjective Quality Assessment of Video Encoded Using H.265 Standard

The H.265 standard [37], also known as high-efficiency video coding (HEVC), was originally published on 13 April 2013 and is currently the most recent and most efficient video coding system. This standard was created in cooperation between the Video Coding Experts Group (VCEG) and the Moving Picture Experts Group. The H.265 standard ensures the compression of videos in very high resolution (2 K, 4 K, 8 K, etc.) and also allows the use of images with increasingly higher resolutions on mobile devices. The H.265 standard offers up to twice the compression compared to H.264. Video compression is based on motion prediction; that is, when there are no changes to a pixel, the codec references that pixel instead of reproducing it. The motion prediction and compensation procedure have also been improved. Another improvement is the enlargement of the macroblock from 16 × 16 pixels (H.264) to 64 × 64 pixels, which is especially important in high-definition movies. The quality degradation of the video signal encoded in the H265 standard was evaluated in a group of 35 people, under both home and laboratory conditions. The results of the measurements taken at home are presented in Table 4 and graphically represented in Figure 8, with the laboratory conditions presented in Table 5 and Figure 9. In addition to the mean MOS value, the tables also include the standard deviation (S) and the values of the confidence interval coefficient (δ) calculated in accordance with the ITU BT.500 recommendation [1]. The statistical analysis of the results showed that up to a bit rate of 600 kbps, the resolution does not affect the evaluation of video image quality. In turn, by comparing the quality ratings for 1280 × 720 and 1920 × 1080 resolutions, it can be observed that there is no significant difference in the image quality rating for bit rates up to 900 kbps for home measurements and up to 1000 kbps under laboratory conditions.

Table 4. Mean value of the video quality assessment (MOS) for H.265 codec, standard deviation (S), and confidence interval coefficient (δ) for four resolutions—measurements at home.

Bit Rate	**640 × 360**			**858 × 480**			**1280 × 720**			**1920 × 1080**		
(kbps)	**MOS**	**S**	**δ**	**MOS**	**S**	**δ**	**MOS**	**S**	**δ**	**MOS**	**S**	**δ**
300	1.09	0.34	0.16	1.09	0.33	0.11	1.09	0.28	0.10	1.12	0.38	0.13
400	1.33	0.56	0.22	1.35	0.50	0.17	1.48	0.59	0.20	1.42	0.65	0.22
500	1.73	0.79	0.33	1.74	0.85	0.29	1.76	0.74	0.25	1.79	0.73	0.24
600	1.97	0.73	0.26	1.97	0.71	0.24	2.16	0.62	0.21	2.18	0.71	0.24
700	2.09	0.72	0.26	2.23	0.77	0.27	2.42	0.67	0.24	2.50	0.76	0.26
800	2.33	0.66	0.22	2.55	0.82	0.28	2.69	0.65	0.22	2.82	0.72	0.25
900	2.47	0.56	0.22	2.76	0.74	0.25	2.94	0.55	0.19	3.06	0.55	0.19
1000	2.56	0.59	0.27	2.94	0.73	0.25	3.13	0.29	0.10	3.35	0.57	0.19
1500	2.75	0.72	0.23	3.18	0.58	0.19	3.44	0.51	0.18	3.79	0.57	0.19
2000	2.93	0.84	0.22	3.35	0.56	0.19	3.64	0.49	0.17	4.03	0.57	0.19
2500	3.06	0.86	0.33	3.58	0.58	0.20	3.85	0.53	0.18	4.26	0.68	0.23
3000	3.24	0.81	0.21	3.68	0.60	0.20	4.00	0.60	0.21	4.44	0.65	0.22
3500	3.33	0.75	0.16	3.82	0.85	0.29	4.18	0.54	0.18	4.53	0.58	0.19
4000	3.44	0.72	0.15	3.88	0.61	0.21	4.25	0.58	0.20	4.65	0.46	0.15
4500	3.52	0.78	0.00	4.03	0.62	0.21	4.31	0.60	0.21	4.71	0.41	0.14
5000	3.63	0.65	0.20	4.09	0.63	0.22	4.39	0.65	0.23	4.79	0.41	0.14
5500	3.69	0.63	0.16	4.18	0.76	0.26	4.53	0.58	0.20	4.85	0.37	0.13
6000	3.84	0.80	0.19	4.21	0.78	0.27	4.59	0.51	0.18	4.88	0.37	0.13

Above these bit rates, the video quality is clearly resolution dependent. For a video signal with a resolution of 1920 × 1080, the MOS rating exceeds the value of 4.0 for the bit rate starting from 2000 kbps for home measurements and approximately 1500 kbps for laboratory measurements. On the other hand, for the resolution of 1280 × 720, the MOS value = 4.0 is reached for a bit rate of 3000 kbps for home measurements, and for laboratory measurements for a bit rate of 2000 kbps. Level 4.0 was also exceeded for a resolution of 858 × 480 with a bit rate of at least 4500 kbps for the home measurements and 2500 kbps for laboratory measurements. A video signal with a resolution of 640 × 360 under home conditions does not reach MOS = 4.0, while under laboratory conditions, starting at 4500 kbps, the MOS reaches a value of 4.0.

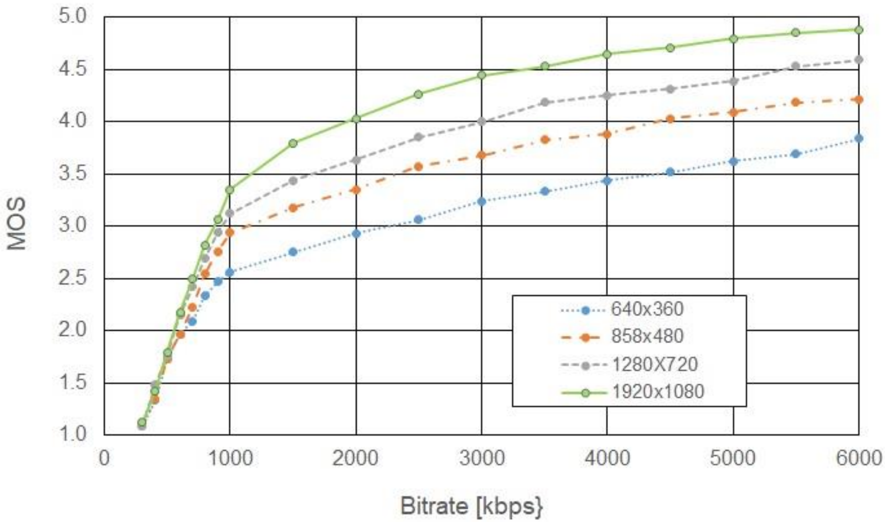

Figure 8. Results of the subjective quality assessment (MOS) for the H.265-encoded video as a function of bit rate for different spatial resolutions—measurements at home.

Table 5. Mean value of the video quality assessment (MOS) for H.265 codec, standard deviation (S), and confidence interval coefficient (δ) for four resolutions—measurements in laboratory.

Bit Rate	**640 × 360**			**858 × 480**			**1280 × 720**			**1920 × 1080**		
(kbps)	**MOS**	**S**	**δ**	**MOS**	**S**	**δ**	**MOS**	**S**	**δ**	**MOS**	**S**	**δ**
300	1.15	0.37	0.16	1.22	0.42	0.17	1.09	0.29	0.12	1.13	0.34	0.14
400	1.55	0.51	0.22	1.43	0.51	0.22	1.61	0.50	0.20	1.65	0.49	0.20
500	1.85	0.75	0.33	1.96	0.71	0.29	2.09	0.42	0.17	2.04	0.64	0.26
600	2.15	0.59	0.26	2.26	0.54	0.22	2.48	0.59	0.24	2.48	0.51	0.21
700	2.35	0.59	0.26	2.57	0.66	0.27	2.87	0.34	0.14	2.83	0.58	0.24
800	2.50	0.51	0.22	2.87	0.55	0.22	3.09	0.42	0.17	3.13	0.76	0.31
900	2.60	0.50	0.22	3.05	0.49	0.20	3.26	0.45	0.18	3.35	0.49	0.20
1000	2.80	0.62	0.27	3.17	0.49	0.20	3.43	0.51	0.21	3.48	0.59	0.24
1500	3.20	0.52	0.23	3.57	0.51	0.21	3.78	0.42	0.17	3.91	0.51	0.21
2000	3.40	0.50	0.22	3.91	0.60	0.24	4.09	0.67	0.27	4.22	0.52	0.21
2500	3.55	0.76	0.33	4.04	0.47	0.19	4.30	0.56	0.23	4.43	0.51	0.21
3000	3.70	0.47	0.21	4.17	0.39	0.16	4.43	0.59	0.24	4.57	0.51	0.21
3500	3.80	0.41	0.18	4.26	0.45	0.18	4.48	0.59	0.24	4.70	0.47	0.19
4000	3.89	0.32	0.14	4.30	0.47	0.19	4.57	0.51	0.21	4.78	0.42	0.17
4500	4.00	0.00	0.00	4.35	0.49	0.20	4.61	0.50	0.20	4.83	0.39	0.16
5000	4.05	0.39	0.17	4.39	0.50	0.20	4.70	0.47	0.19	4.87	0.34	0.14
5500	4.10	0.31	0.13	4.39	0.50	0.20	4.74	0.45	0.18	4.91	0.29	0.12
6000	4.15	0.37	0.16	4.41	0.50	0.21	4.78	0.42	0.17	4.91	0.29	0.12

Compared to the H.264-encoding standard, much higher MOS rating values are observed for the H.265 standard. Comparing the results of the MOS evaluation obtained under laboratory and home conditions, it can be seen that the viewers rated the video image presented under laboratory conditions more highly; Table 6 and Figure 10 show the difference of ΔMOS in the evaluation of video quality obtained for measurements made under laboratory and home conditions according to Formula (1).

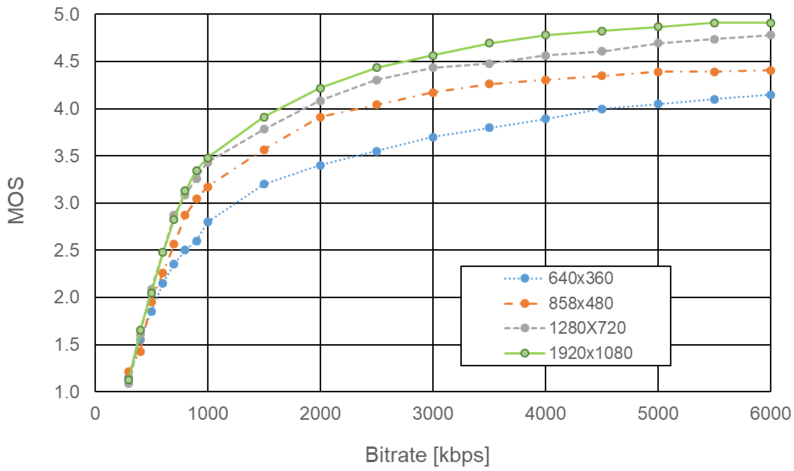

Figure 9. Results of the subjective quality assessment (MOS) for the H.265-encoded video as a function of bit rate for different spatial resolutions—measurements in the laboratory.

Table 6. Value of the difference in ΔMOS (for H.265 codec) between the results obtained under laboratory and home conditions.

Bit Rate	**ΔMOS**			
(kbps)	**640 × 360**	**858 × 480**	**1280 × 720**	**1920 × 1080**
300	0.06	0.13	0.00	0.01
400	0.22	0.08	0.12	0.23
500	0.12	0.22	0.33	0.25
600	0.18	0.29	0.32	0.30
700	0.26	0.34	0.45	0.33
800	0.17	0.32	0.40	0.31
900	0.13	0.29	0.32	0.29
1000	0.24	0.23	0.31	0.13
1500	0.45	0.39	0.35	0.12
2000	0.47	0.56	0.45	0.19
2500	0.49	0.47	0.46	0.17
3000	0.46	0.50	0.43	0.12
3500	0.47	0.44	0.30	0.17
4000	0.46	0.43	0.32	0.14
4500	0.48	0.32	0.30	0.12
5000	0.43	0.30	0.31	0.08
5500	0.41	0.21	0.21	0.06
6000	0.31	0.20	0.19	0.03

The statistical analysis of the results obtained using the *t*-test showed that at the confidence level of $\alpha = 0.05$, there is no basis to accept the hypothesis of the identity of the results obtained under laboratory and home conditions. The *t*-test values obtained with the Statistica tool for each resolution are as follows:

- 640 × 360; $t = 9.1 > t\ \alpha = 2.1$, at $\alpha = 0.05$;
- 858 × 480; $t = 10.4 > t\ \alpha = 2.1$, at $\alpha = 0.05$;
- 1280 × 720; $t = 10.9 > t\ \alpha = 2.1$, at $\alpha = 0.05$;
- 1920 × 1080; $t = 7.3 > t\ \alpha = 2.1$, at $\alpha = 0.05$.

It can be concluded that the differences between the MOS values obtained under the laboratory and home conditions are significant.

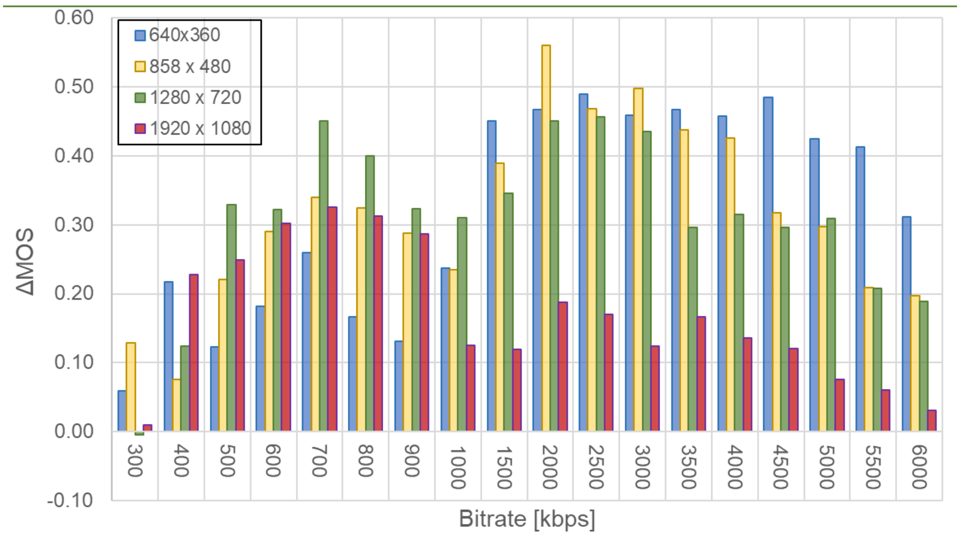

Figure 10. Difference between the results of the subjective evaluation of the H.265-encoded video (ΔMOS), conducted in the laboratory and at home, as a function of bit rate for different resolutions.

3.3. Subjective Quality Assessment of Video Encoded Using VP9 Standard

The VP9 standard, developed by Google, was the last evaluated coding technique. The VP9 codec is used, among others, on YouTube. The VP9 codec is based on an open-source license, uses the Webm container, and is basically MKV (the H.264 and H.265 codecs use the MP4 container) [40]. The degradation of the quality of the video signal encoded in the VP9 standard was assessed in a group of 30 people at home and 40 people in a laboratory. The results of the measurements made under home conditions are presented in Table 7 and graphically in Figure 11, with the results of the laboratory conditions in Table 8 and Figure 12. In addition to the MOS mean value, the tables also include the standard deviation values (S), as well as the values of the confidence interval coefficient (δ) calculated according to the ITU BT.500 recommendation [1]. The statistical analysis of the results showed that up to a bit rate of 2000 kbps, there is no difference in the assessment of image quality with resolutions of 1920 × 1080 and 1280 × 720 in home measurements; for higher speeds, slight differences can be observed in favor of the image with higher resolution. However, the differences in the quality assessment are within the designated confidence interval. For both resolutions, the MOS value of 4.0 is exceeded at 3000 kbps. The quality assessment made at home for other resolutions is comparable to the H.265 standard, which is probably related to the young people's habits, as this standard is very popular, among others, on YouTube. In turn, the statistical analysis of the results obtained in the laboratory measurements showed that up to a bit rate of 1000 kbps, there is no difference in the assessment of image quality for all of the resolutions assessed. Above this bit rate, the video quality slightly depends on the resolution and, as expected, the video signal with a resolution of 1920 × 1080 is rated the highest, for which MOS = 4 was already achieved at a bit rate of 2000 kbps. An MOS value of 4.0 was obtained for 1280 × 720 at a bit rate of 2500 kbps and for 858 × 480 at a bit rate of 3000 kbps. The smallest resolution, i.e., 640 × 360, achieves the worst MOS values, but starting from 4500 kbps, the quality rating reaches the level of 4.0, just like in home measurements.

Table 7. Mean value of the video quality assessment (MOS) for VP9 codec, standard deviation (S), and confidence interval coefficient (δ) for four resolutions—measurements at home.

Bit Rate	640 × 360			858 × 480			1280 × 720			1920 × 1080		
(kbps)	MOS	S	δ	MOS	S	δ	MOS	S	δ	MOS	S	δ
300	1.12	0.35	0.14	1.24	0.46	0.18	1.26	0.45	0.13	1.29	0.48	0.19
400	1.48	0.51	0.20	1.52	0.51	0.20	1.69	0.56	0.17	1.62	0.76	0.29
500	1.76	0.43	0.17	1.80	0.71	0.28	2.07	0.64	0.19	2.00	0.86	0.32
600	1.92	0.56	0.22	1.96	0.71	0.28	2.41	0.55	0.17	2.25	0.75	0.28
700	2.09	0.57	0.23	2.04	0.74	0.30	2.71	0.55	0.17	2.43	0.92	0.34
800	2.30	0.62	0.24	2.29	0.58	0.23	2.93	0.71	0.22	2.64	0.85	0.32
900	2.38	0.66	0.25	2.46	0.60	0.24	3.12	0.50	0.15	2.85	0.80	0.30
1000	2.48	0.51	0.20	2.60	0.50	0.20	3.26	0.45	0.13	3.00	0.76	0.28
1500	2.81	0.74	0.28	2.96	0.71	0.27	3.64	0.48	0.15	3.37	0.81	0.31
2000	2.96	0.85	0.33	3.19	0.72	0.28	3.90	0.43	0.13	3.71	0.56	0.21
2500	3.07	0.72	0.27	3.35	0.70	0.27	4.07	0.51	0.16	3.93	0.57	0.21
3000	3.19	0.72	0.27	3.50	0.59	0.23	4.29	0.60	0.18	4.14	0.60	0.22
3500	3.30	0.64	0.24	3.64	0.66	0.26	4.43	0.55	0.17	4.36	0.56	0.21
4000	3.44	0.59	0.22	3.77	0.62	0.24	4.52	0.55	0.17	4.46	0.51	0.19
4500	3.52	0.51	0.19	3.85	0.58	0.22	4.57	0.55	0.17	4.61	0.51	0.19
5000	3.63	0.65	0.24	3.96	0.71	0.27	4.64	0.48	0.15	4.67	0.49	0.19
5500	3.70	0.62	0.24	4.04	0.71	0.27	4.69	0.47	0.14	4.70	0.48	0.18
6000	3.74	0.61	0.23	4.15	0.63	0.24	4.74	0.45	0.13	4.75	0.46	0.17

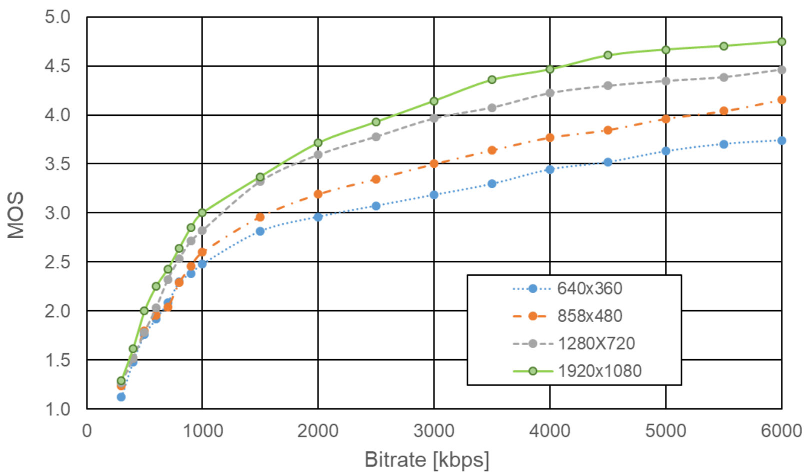

Figure 11. Results of the subjective quality assessment (MOS) for the VP9-encoded video as a function of bit rate for different spatial resolutions—measurements at home.

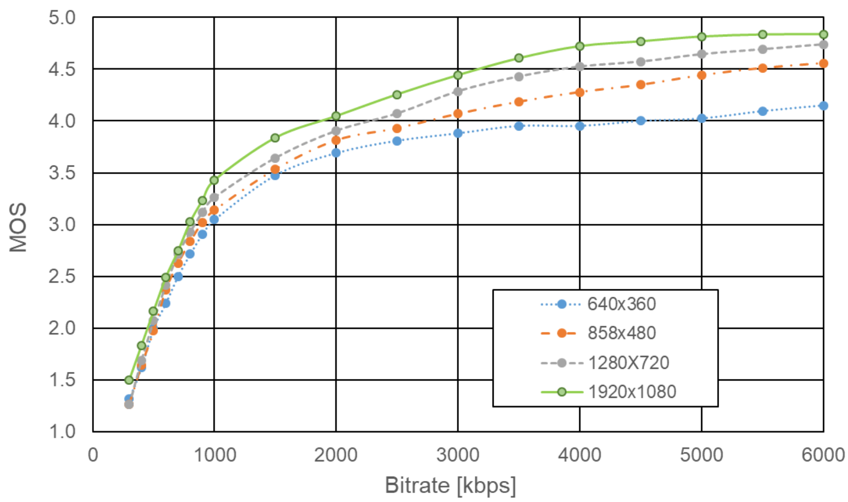

Figure 12. Results of the subjective quality assessment (MOS) for the VP9-encoded video as a function of bit rate for different spatial resolutions—measurements in laboratory.

Comparing the results of the MOS evaluation obtained in the laboratory and home conditions, it can be seen that the viewers rated the video image presented under laboratory conditions more highly; Table 9 and Figure 13 show the difference in ΔMOS in the evaluation of video quality obtained for measurements made under laboratory and home conditions according to Formula (1).

Table 8. Mean value of the video quality assessment (MOS) for VP9 codec, standard deviation (S), and confidence interval coefficient (δ) for four resolutions—measurements in laboratory.

Bit Rate	640 × 360			858 × 480			1280 × 720			1920 × 1080		
(kbps)	MOS	S	δ	MOS	S	δ	MOS	S	δ	MOS	S	δ
300	1.32	0.47	0.14	1.27	0.45	0.14	1.26	0.45	0.13	1.50	0.55	0.17
400	1.62	0.58	0.18	1.64	0.58	0.17	1.69	0.56	0.17	1.83	0.66	0.20
500	2.00	0.44	0.13	1.98	0.51	0.15	2.07	0.64	0.19	2.17	0.61	0.20
600	2.24	0.48	0.15	2.37	0.62	0.18	2.41	0.55	0.17	2.49	0.63	0.19
700	2.50	0.51	0.15	2.63	0.49	0.15	2.71	0.55	0.17	2.74	0.76	0.23
800	2.71	0.46	0.14	2.84	0.43	0.13	2.93	0.71	0.22	3.02	0.47	0.14
900	2.90	0.30	0.09	3.02	0.34	0.10	3.12	0.50	0.15	3.23	0.43	0.13
1000	3.05	0.44	0.13	3.14	0.41	0.12	3.26	0.45	0.13	3.43	0.50	0.15
1500	3.48	0.51	0.15	3.53	0.50	0.15	3.64	0.48	0.15	3.84	0.48	0.14
2000	3.69	0.56	0.17	3.81	0.45	0.13	3.90	0.43	0.13	4.05	0.43	0.13
2500	3.81	0.45	0.14	3.93	0.40	0.12	4.07	0.51	0.16	4.26	0.44	0.13
3000	3.88	0.33	0.10	4.07	0.40	0.12	4.29	0.60	0.18	4.44	0.50	0.15
3500	3.95	0.38	0.11	4.19	0.50	0.15	4.43	0.55	0.17	4.60	0.49	0.15
4000	3.95	0.44	0.13	4.28	0.50	0.15	4.52	0.55	0.17	4.72	0.45	0.14
4500	4.00	0.58	0.18	4.35	0.53	0.16	4.57	0.55	0.17	4.77	0.43	0.13
5000	4.05	0.44	0.14	4.44	0.50	0.15	4.64	0.48	0.15	4.81	0.39	0.12
5500	4.10	0.43	0.13	4.51	0.51	0.15	4.69	0.47	0.14	4.83	0.38	0.11
6000	4.15	0.48	0.15	4.56	0.50	0.15	4.74	0.45	0.13	4.84	0.37	0.11

Table 9. Value of the difference in ΔMOS (for VP9 codec) between the results obtained under laboratory and home conditions.

Bit Rate	ΔMOS			
(kbps)	640 × 360	858 × 480	1280 × 720	1920 × 1080
300	0.20	0.03	−0.01	0.21
400	0.14	0.12	0.17	0.22
500	0.24	0.18	0.29	0.17
600	0.32	0.41	0.38	0.24
700	0.41	0.59	0.39	0.32
800	0.42	0.55	0.39	0.38
900	0.52	0.56	0.40	0.38
1000	0.57	0.54	0.44	0.43
1500	0.66	0.57	0.32	0.47
2000	0.73	0.62	0.31	0.33
2500	0.74	0.58	0.29	0.33
3000	0.70	0.57	0.32	0.30
3500	0.66	0.55	0.35	0.25
4000	0.51	0.51	0.30	0.26
4500	0.48	0.50	0.28	0.16
5000	0.42	0.48	0.30	0.15
5500	0.39	0.47	0.31	0.13
6000	0.41	0.40	0.28	0.09

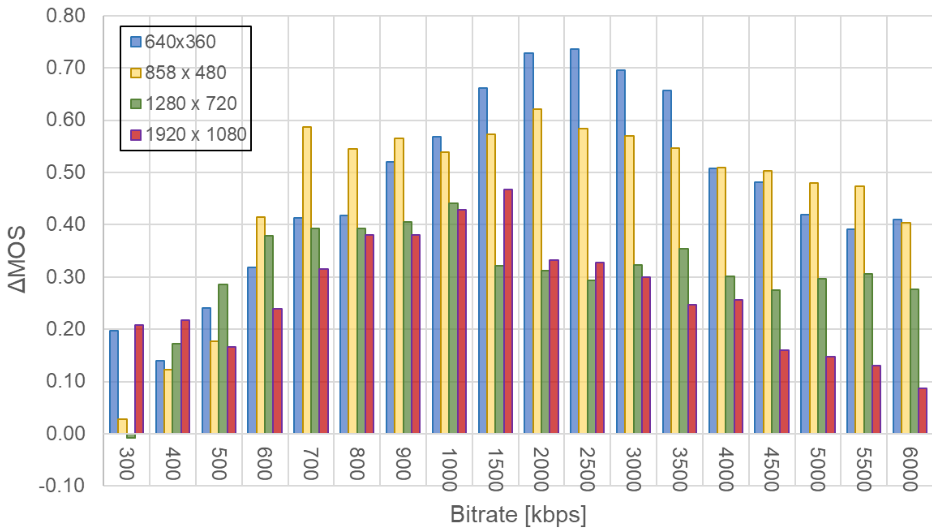

Figure 13. Difference between the results of the subjective evaluation of the VP9-encoded video (ΔMOS), conducted in the laboratory and at home, as a function of bit rate for different resolutions.

The statistical analysis of the results obtained using the *t*-test showed that at the confidence level of $\alpha = 0.05$, there is no basis to accept the hypothesis of the identity of the results obtained under laboratory and home conditions. The *t*-test values obtained with the Statistica tool for each resolution are as follows:

- 640 × 360; t = 11.1 > t α = 2.1, at $\alpha = 0.05$;
- 858 × 480; t = 11.3 > t α = 2.1, at $\alpha = 0.05$;
- 1280 × 720; t = 13.0 > t α = 2.1, at $\alpha = 0.05$;
- 1920 × 1080; t = 10.5 > t α = 2.1, at $\alpha = 0.05$.

It can be concluded that the differences between the MOS values obtained under the laboratory and home conditions are significant.

3.4. Objective Quality Assessment of Video Encoded Using H.264, H.265, and VP9 Standards

The results of the objective video quality assessment are presented using three metrics: PSNR, SSIM, and VMAF (see Figures 14–16).

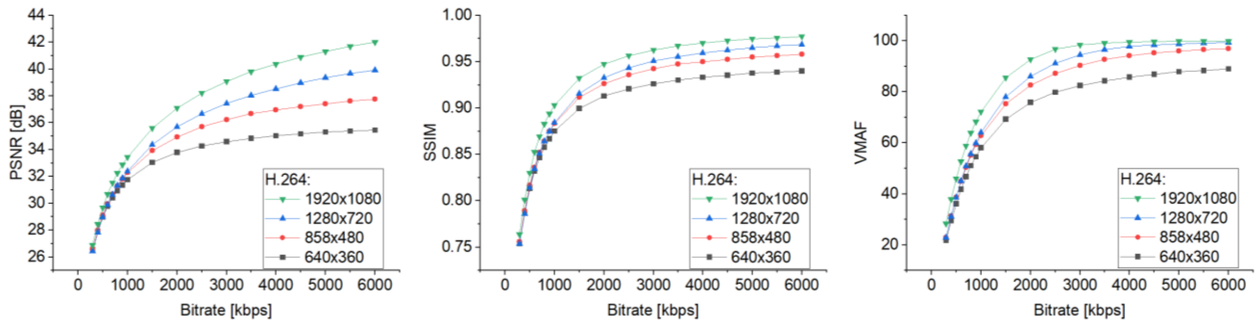

Figure 14. Relationship of the objective assessment of video quality encoded in the H.264 standard vs. bit rate for the resolutions 640 × 360, 858 × 480, 1280 × 720, and 1920 × 1080.

It can be noted that, just like in the case of subjective quality assessment, the objective video quality results are directly proportional to the used coding bit rate, which is valid for all presented metrics and video codecs. The most significant changes in quality are observed for low bit rates, while for higher bit rates, the quality changes seem to be very small or imperceptible. The results are also consistent with those presented in the literature, where the H.265 and VP9 codecs are more efficient than the H.264 codec. A very important issue here is the problem of the spatial resolutions of the examined videos. Here, each set of video samples of a specific resolution was compared (double-stimulus method) with a reference footage of proper resolution, i.e., 360p reference with 360p test sample, 480p reference with 480p test sample, etc. This resulted in higher quality values for videos with higher spatial resolution, which was consistent with the results of the subjective assessment. The correlation coefficients between these objective results and subjective quality assessment scores in the laboratory and in users' homes for each codec and video spatial resolution were determined and are presented in Tables 10–12.

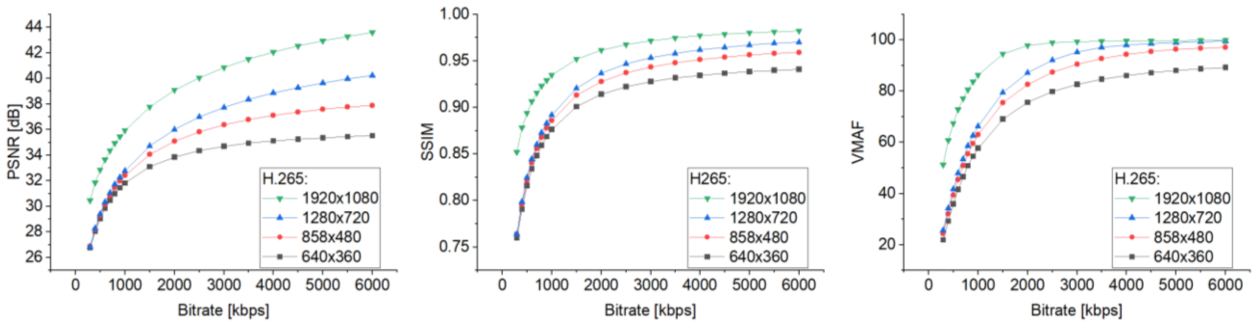

Figure 15. Relationship of the objective assessment of video quality encoded in the H.265 standard vs. bit rate for the resolutions 640 × 360, 858 × 480, 1280 × 720, and 1920 × 1080.

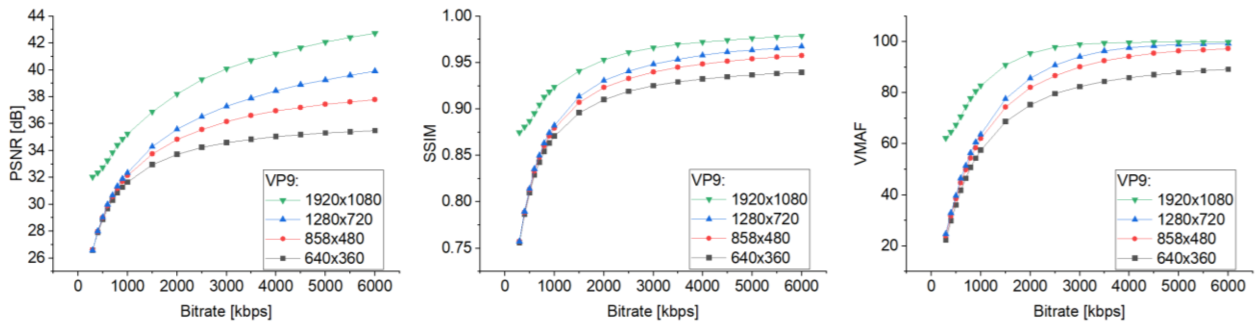

Figure 16. Relationship of the objective assessment of video quality encoded in the VP9 standard vs. bit rate for the resolutions 640 × 360, 858 × 480, 1280 × 720, and 1920 × 1080.

Table 10. Correlations between QoS and QoE values (in the lab and home) for H.264 encoded video.

		QoS vs. QoE Correlations							
		Lab				**Home**			
		360p	**480p**	**720p**	**1080p**	**360p**	**480p**	**720p**	**1080p**
PSNR	360p	0.981	0.996	0.995	0.986	0.986	0.988	0.995	0.985
	480p	0.968	0.989	0.990	0.987	0.980	0.989	0.998	0.994
	720p	0.951	0.978	0.981	0.982	0.970	0.986	0.995	0.996
	1080p	0.949	0.976	0.979	0.981	0.968	0.985	0.994	0.996
SSIM	360p	0.989	0.997	0.995	0.977	0.988	0.983	0.987	0.970
	480p	0.987	0.997	0.996	0.981	0.988	0.985	0.990	0.975
	720p	0.985	0.997	0.996	0.984	0.988	0.987	0.993	0.980
	1080p	0.989	0.997	0.995	0.978	0.988	0.982	0.987	0.970
VMAF	360p	0.974	0.992	0.994	0.992	0.983	0.988	0.998	0.992
	480p	0.973	0.992	0.993	0.992	0.982	0.987	0.998	0.993
	720p	0.971	0.990	0.992	0.993	0.979	0.984	0.996	0.992
	1080p	0.981	0.994	0.994	0.990	0.982	0.979	0.989	0.980

Table 11. Correlations between QoS and QoE values (in the lab and home) for H.265 encoded video.

		QoS vs. QoE Correlations							
		Lab				Home			
		360p	480p	720p	1080p	360p	480p	720p	1080p
PSNR	360p	0.998	0.997	0.997	0.998	0.991	0.995	0.996	0.997
	480p	0.999	0.991	0.990	0.993	0.993	0.994	0.995	0.996
	720p	0.995	0.981	0.980	0.983	0.991	0.989	0.991	0.989
	1080p	0.991	0.975	0.973	0.977	0.989	0.986	0.987	0.985
SSIM	360p	0.993	0.997	0.999	0.998	0.986	0.991	0.992	0.993
	480p	0.995	0.998	0.999	0.999	0.988	0.993	0.994	0.995
	720p	0.995	0.998	0.999	0.999	0.988	0.993	0.994	0.995
	1080p	0.994	0.996	0.998	0.998	0.987	0.991	0.993	0.993
VMAF	360p	0.998	0.995	0.992	0.995	0.989	0.994	0.995	0.998
	480p	0.997	0.995	0.992	0.995	0.988	0.993	0.995	0.997
	720p	0.995	0.996	0.992	0.995	0.985	0.991	0.993	0.996
	1080p	0.976	0.989	0.991	0.989	0.965	0.976	0.977	0.981

Table 12. Correlations between QoS and QoE values (in the lab and home) for VP9 encoded video.

		QoS vs. QoE Correlations							
		Lab				Home			
		360p	480p	720p	1080p	360p	480p	720p	1080p
PSNR	360p	0.997	0.998	0.999	0.998	0.994	0.990	0.996	0.993
	480p	0.990	0.996	0.997	0.997	0.997	0.997	0.999	0.999
	720p	0.980	0.989	0.990	0.991	0.995	0.999	0.997	0.999
	1080p	0.951	0.966	0.968	0.972	0.978	0.989	0.984	0.987
SSIM	360p	0.998	0.997	0.997	0.995	0.988	0.981	0.989	0.985
	480p	0.998	0.998	0.998	0.996	0.991	0.984	0.991	0.988
	720p	0.998	0.998	0.998	0.997	0.991	0.986	0.993	0.990
	1080p	0.988	0.991	0.992	0.995	0.990	0.994	0.999	0.995
VMAF	360p	0.995	0.996	0.997	0.998	0.993	0.993	0.998	0.995
	480p	0.994	0.995	0.996	0.998	0.992	0.993	0.998	0.996
	720p	0.994	0.994	0.995	0.997	0.989	0.990	0.997	0.993
	1080p	0.993	0.988	0.987	0.991	0.978	0.980	0.991	0.983

All tests and calculated correlations were conducted for the selected video resolutions and a limited number of coding bit rates (i.e., 18 coding bit rates for each video sample of a specific resolution). To validate these results and check how much of the whole population of different cases is well described by this research, a determination coefficient (R^2) was calculated for each previously determined correlation.

Taking into account each codec, it can be stated that the determination coefficients fluctuated as follows:

1. For the H.264 codec: from 0.9 to 0.996;
2. For the H.265 codec: from 0.931 to 0.998;
3. For the VP9 codec: from 0.905 to 0.998.

This means that the obtained results of the correlations well describe 90 to 99 percent of the whole population. This leads to the conclusion that the obtained correlations are very strong and that they are representative for a population that can be much wider than the video set that was used during the research.

4. Conclusions

The authors presented the problem of subjective quality assessment conducted in different environments. Most papers and formal regulations recommend performing such tests in a laboratory under special circumstances. It is understandable that test conditions must be strictly determined, especially when the procedure must be repeatable and should give representative results that are comparable with those of other laboratories. However, in the case of watching the video at home, the environment may not meet the laboratory conditions described in formal recommendations. This may cause the quality experienced by the home user to differ from the quality measured in the laboratory. The results of our investigations confirmed these assumptions and showed statistically significant differences. This implies the need to separate these two types of environments and to conduct the tests in both depending on their purpose. The second part of the research was devoted to objective video quality evaluation and identifying the relationships between their results and the results of the subjective assessment conducted in different environments. The authors observed very high correlations between all three sets of results, i.e., objective, subjective in the laboratory, and subjective at home. The very high determination coefficients imply that the results obtained from testing a limited number of video samples may produce conclusions that can be generalized to the entire population. Obviously, here, the QoS/QoE models can be made, but their parameters must be determined separately for laboratory and home environments. Searching for better quality models for other environments (different from laboratory) may help to better fit the video delivered to its recipients.

Author Contributions: Conceptualization and methodology, J.K. and S.B.; objective quality assessment, J.K.; subjective quality assessment, S.B. with M.Ł.'s support; writing—original draft preparation, J.K. and S.B.; visualization, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data sharing is not applicable to this article.

Acknowledgments: The paper presents the results of the statutory research carried out at Wroclaw University of Science and Technology. The authors would like to thank Wroclaw Centre for Networking and Supercomputing for providing the computing resources that were used for the digital processing of the tested video samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. *ITU-R BT 500-14*; Methodologies for the Subjective Assessment of the Quality of Television Images. ITU: Geneva, Switzerland, 2020.
2. Fela, R.F.; Zacharov, N.; Forchhammer, S. Comparison of Full Factorial and Optimal Experimental Design for Perceptual Evaluation of Audiovisual Quality. *J. Audio Eng. Soc.* **2023**, *71*, 4–19. [CrossRef]
3. Barman, N.; Martini, M.G. QoE Modeling for HTTP Adaptive Video Streaming–A Survey and Open Challenges. *IEEE Access* **2019**, *7*, 30831–30859. [CrossRef]
4. Wang, Z.; Bovik, A.C. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **2009**, *26*, 98–117. [CrossRef]
5. Huynh-Thu, Q.; Ghanbari, M. The accuracy of PSNR in predicting video quality for different video scenes and frame rates. *Telecommun. Syst.* **2012**, *49*, 35–48. [CrossRef]
6. Klink, J.; Uhl, T. Video Quality Assessment: Some Remarks on Selected Objective Metrics. In Proceedings of the 2020 28th International Conference Software, Telecommun. Comput. Networks, SoftCOM, Split, Croatia, 17–19 September 2020.
7. Vranjes, M.; Rimac-Drlje, S.; Zagar, D. Objective video quality metrics. In Proceedings of the ELMAR 2007, Zadar, Croatia, 12–14 September 200; pp. 45–49. [CrossRef]
8. Kotevski, Z.; Mitrevski, P. Performance Assessment of Metrics for Video Quality Estimation. In Proceedings of the International Scientific Conference on Information, Communication and Energy Systems and Technologies, Macedonia, Greece, 23–26 June 2010; pp. 693–696.
9. Chikkerur, S.; Sundaram, V.; Reisslein, M.; Karam, L.J. Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE Trans. Broadcast.* **2011**, *57*, 165–182. [CrossRef]

10. Akramullah, S.; Akramullah, S. Video quality metrics. In *Digital Video Concepts, Methods, and Metrics*; Apress: New York, NY, USA, 2014; pp. 101–160.
11. Chen, Y.; Wu, K.; Zhang, Q. From QoS to QoE: A Tutorial on Video Quality Assessment. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 1126–1165. [CrossRef]
12. Hanhart, P.; Korshunov, P.; Ebrahimi, T. Benchmarking of quality metrics on ultra-high definition video sequences. In Proceedings of the 2013 18th International Conference on Digital Signal Processing (DSP), Santorini, Greece, 1–3 July 2013; pp. 1–8.
13. Hanhart, P.; Bernardo, M.V.; Pereira, M.; Pinheiro, A.M.G.; Ebrahimi, T. Benchmarking of objective quality metrics for HDR image quality assessment. *EURASIP J. Image Video Process.* **2015**, *2015*, 39. [CrossRef]
14. Klink, J. A Method of Codec Comparison and Selection for Good Quality Video Transmission Over Limited-Bandwidth Networks. *Sensors* **2021**, *21*, 4589. [CrossRef]
15. Barman, N.; Martini, M.G. H. 264/MPEG-AVC, H. 265/MPEG-HEVC and VP9 codec comparison for live gaming video streaming. In Proceedings of the 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017; pp. 1–6.
16. You, J.; Reiter, U.; Hannuksela, M.M.; Gabbouj, M.; Perkis, A. Perceptual-based quality assessment for audio–visual services: A survey. *Signal Process. Image Commun.* **2010**, *25*, 482–501. [CrossRef]
17. Akhtar, Z.; Siddique, K.; Rattani, A.; Lutfi, S.L.; Falk, T.H. Why is Multimedia Quality of Experience Assessment a Challenging Problem? *IEEE Access* **2017**, *7*, 117897–117915. [CrossRef]
18. Rassool, R. VMAF reproducibility: Validating a perceptual practical video quality metric. In Proceedings of the 2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Cagliari, Italy, 7–9 June 2017; pp. 1–2.
19. Moldovan, A.-N.; Ghergulescu, I.; Muntean, C.H. VQAMap: A Novel Mechanism for Mapping Objective Video Quality Metrics to Subjective MOS Scale. *IEEE Trans. Broadcast.* **2016**, *62*, 610–627. [CrossRef]
20. Bentaleb, A.; Taani, B.; Begen, A.C.; Timmerer, C.; Zimmermann, R. A Survey on Bitrate Adaptation Schemes for Streaming Media Over HTTP. *IEEE Commun. Surv. Tutor.* **2018**, *21*, 562–585. [CrossRef]
21. Sani, Y.; Mauthe, A.; Edwards, C. Adaptive Bitrate Selection: A Survey. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 2985–3014. [CrossRef]
22. Zabrovskiy, A.; Feldmann, C.; Timmerer, C. Multi-codec DASH dataset. In Proceedings of the 9th ACM Multimedia Systems Conference, Amsterdam, The Netherlands, 12–15 June 2018; pp. 438–443.
23. Tanchenko, A. Visual-PSNR measure of image quality. *J. Vis. Commun. Image Represent.* **2014**, *25*, 874–878. [CrossRef]
24. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
25. Bampis, C.G.; Li, Z.; Bovik, A.C. Spatiotemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 2256–2270. [CrossRef]
26. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
27. Li, L.; Lin, W.; Wang, X.; Yang, G.; Bahrami, K.; Kot, A.C. No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments. *IEEE Trans. Cybern.* **2015**, *46*, 39–50. [CrossRef]
28. Li, L.; Zhu, H.; Yang, G.; Qian, J. Referenceless Measure of Blocking Artifacts by Tchebichef Kernel Analysis. *IEEE Signal Process. Lett.* **2013**, *21*, 122–125. [CrossRef]
29. Liu, H.; Klomp, N.; Heynderickx, I. A No-Reference Metric for Perceived Ringing Artifacts in Images. *IEEE Trans. Circuits Syst. Video Technol.* **2009**, *20*, 529–539. [CrossRef]
30. Zhu, H.; Li, L.; Wu, J.; Dong, W.; Shi, G. MetaIQA: Deep meta-learning for no-reference image quality assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14 June 2020–19 June 2020; pp. 14143–14152.
31. Kang, L.; Ye, P.; Li, Y.; Doermann, D. Convolutional neural networks for no-reference image quality assessment. In Proceedings of the 2014 IEEE Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1733–1740.
32. Bosse, S.; Maniry, D.; Müller, K.R.; Wiegand, T.; Samek, W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans. Image Process.* **2017**, *27*, 206–219. [CrossRef] [PubMed]
33. Zhang, W.; Ma, K.; Yan, J.; Deng, D.; Wang, Z. Blind Image Quality Assessment Using a Deep Bilinear Convolutional Neural Network. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *30*, 36–47. [CrossRef]
34. Sun, S.; Yu, T.; Xu, J.; Zhou, W.; Chen, Z. GraphIQA: Learning Distortion Graph Representations for Blind Image Quality Assessment. *IEEE Trans. Multimedia* **2022**, *14*, 1–14. [CrossRef]
35. Liu, J.; Zhou, W.; Li, X.; Xu, J.; Chen, Z. LIQA: Lifelong Blind Image Quality Assessment. *IEEE Trans. Multimedia* **2022**, *14*, 1–13. [CrossRef]
36. *ITU-T Rec*; H.264. Audiovisual and Multimedia Systems: Infrastructure of Audiovisual Services-Coding of Moving Video, Advanced Video Coding for Generic Audiovisual Services. International Telecommunication Union: Geneva, Switzerland, 2021.
37. *ITU-T Rec*; H.265. Infrastructure of Audiovisual Services—Coding of Moving Video. High Efficiency Video Coding. International Telecommunication Union: Geneva, Switzerland, 2021.
38. Grange, A.; De Rivaz, P.; Hunt, J. VP9 Bitstream Decoding Process Specification. *WebM Project.* 2016. Available online: http://downloads.webmproject.org.storage.googleapis.com/docs/vp9/vp9-bitstream-specification-v0.6-20160331-draft.pdf (accessed on 25 February 2023).

39. *ITU-T Rec*; P.910. Subjective Video Quality Assessment Methods for Multimedia Applications. International Telecommunication Union: Geneva, Switzerland, 2021.
40. Mukherjee, D.; Bankoski, J.; Grange, A.; Han, J.; Koleszar, J.; Wilkins, P.; Xu, Y.; Bultje, R. The latest open-source video codec VP9—An overview and preliminary results. In Proceedings of the 2013 Picture Coding Symposium, San Jose, CA, USA, 8–11 December 2013; pp. 390–393. [CrossRef]
41. Winkler, S. Video quality measurement standards—Current status and trends. In Proceedings of the 2009 7th International Conference on Information, Communications and Signal Processing (ICICS), Macau, China, 8–10 December 2009; pp. 1–5.
42. Winkler, S. On the properties of subjective ratings in video quality experiments. In Proceedings of the 2009 International Workshop on Quality of Multimedia Experience, Lippstadt, Germany, 5–7 September 2022; pp. 139–144. [CrossRef]
43. *ITU-T Recommendation P.913*; Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment. ITU: Geneva, Switzerland, 2021.
44. Harysandi, D.K.; Oktaviani, R.; Meylani, L.; Vonnisa, M.; Hashiguchi, H.; Shimomai, T.; Aris, N.A.M. International Telecommunication Union-Radiocommunication Sector P. 837-6 and P. 837-7 performance to estimate Indonesian rainfall. *Telkomnika* **2020**, *18*, 2292–2303.
45. *ITU-R BT 709-6*; Parameter Values for the HDTV Standards for Production and International Programme Exchange BT Series Broadcasting Service. ITU: Geneva, Switzerland, 2015.
46. Taha, M.; Ali, A.; Lloret, J.; Gondim, P.R.L.; Canovas, A. An automated model for the assessment of QoE of adaptive video streaming over wireless networks. *Multimedia Tools Appl.* **2021**, *80*, 26833–26854. [CrossRef]
47. Mercat, A.; Viitanen, M.; Vanne, J. UVG dataset: 50/120fps 4K sequences for video codec analysis and development. In Proceedings of the 11th ACM Multimedia Systems Conference, Istanbul, Turkey, 8– 11 June 2020; pp. 297–302.
48. FFmpeg: A Complete, Cross -Platform Solution to Record, Convert and Stream Audio and Video. Available online: https://ffmpeg.org/ (accessed on 25 February 2023).
49. Brachmański, S.; Klink, J. Subjective Assessment of the Quality of Video Sequences by the Young Viewers. In Proceedings of the 30 th International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2022), Split, Croatia, 22–24 September 2022; pp. 1–6.