NTIA TM-27-585

U.S. Department of Commerce
National Telecommunications and Information Administration

Institute for Telecommunication Sciences
Technical Memorandum

JNR-ITSps: Dataset Track 2-2 from ITU-T Study Group 12 Validation of No-Reference Metrics for Camera Capture Impairments, Depicting the Public Safety User Case

Margaret H. Pinson, John Dumke, and Rob Grosso

October 2026
Boulder, Colorado
its.ntia.gov

Institute for Telecommunication Sciences
Technical Memorandum
NTIA TM-27-585
October 2026

JNR-ITSps: Dataset Track 2-2 from ITU-T Study Group 12 Validation of No-Reference Metrics for Camera Capture Impairments, Depicting the Public Safety User Case

Margaret H. Pinson, John Dumke, and Rob Grosso

United States Department of Commerce

National Telecommunications and Information Administration

Institute for Telecommunication Sciences
325 Broadway, Boulder, CO 80305
its.ntia.gov

Approved for public release; distribution is unlimited.

This publication is available free of charge from:
https://doi.org/10.70220/nqwtpz2l

About the Institute for Telecommunication Sciences

The Institute for Telecommunication Sciences (ITS) is the research and engineering laboratory of the National Telecommunications and Information Administration (NTIA), an agency of the U.S. Department of Commerce (DoC). ITS manages the telecommunications technology research programs of NTIA and the DoC-owned Table Mountain Radio Quiet Zone. ITS works closely with other NTIA Line Offices to support Administration and Agency needs.

The mission of ITS is to ADVANCE innovation in communications technologies, INFORM spectrum and communications policy for the benefit of all stakeholders, and INVESTIGATE our Nation's most pressing telecommunications challenges through research that employees are proud to deliver.

ITS publishes fundamental research communications as part of its Technology Transfer efforts. We are committed to ensuring that our publications are substantive, technically sound, accurate, and clear. A rigorous peer review process, documented in the "ITS Publications Handbook Volume I: Policies (Third Edition)," safeguards the scientific integrity of ITS technical publications, journal articles, and conference papers. The principles for technical peer review of manuscripts guide parallel peer review processes for the publication of software and datasets.

ITS reports authored by NTIA employees are subject to 17 U.S.C. §105 and are U.S. Government works that are generally not subject to copyright protection in the United States. ITS publications, including software and datasets, are freely and openly available at its.ntia.gov or github.com/NTIA.

Principal NTIA Formal Publication Series Published by ITS

NTIA Technical Report (TR): Important contributions to existing knowledge of less breadth than a monograph, such as results of completed projects and major activities.

NTIA Technical Memorandum (TM): Technical information typically of less breadth than an NTIA Technical Report. The series includes data, preliminary project results, and information for a specific, limited audience.

NTIA Special Publication (SP): Conference proceedings; bibliographies; course and instructional materials; or major scientific studies mandated by Congress.

NTIA Monograph (MG): A scholarly, professionally oriented publication dealing with state-of-the-art research, or an authoritative treatment of a broad area. Expected to have long-lasting value.

NTIA Handbook (HB): Information pertaining to technical procedures; reference and data guides; and formal user's manuals that are expected to be pertinent for a long time.

For more information about NTIA fundamental research publications, contact the ITS Publications Office at 325 Broadway, Boulder, CO, 80305 Tel. 303-497-3572 or email ITSinfo@ntia.gov.

Disclaimer

Certain products, technologies, and corporations are mentioned in this report to describe aspects of the ways that video systems are used at present or may be used in the future. The mention of such entities should not be construed as any endorsement, approval, recommendation, prediction of success, or that they are in any way superior to or more noteworthy than similar entities that were not mentioned.

Contents

About the Institute for Telecommunication Sciences ii
Disclaimer iii
Figures v
Tables vi
Preface vii
Glossary of Terms viii
Executive Summary x
1. Introduction 2
2. SRC Videos 3
3. Transcoding Impairments 5
4. Subjective Testing 9
5. Conclusions 11
6. References 12
Appendix A. Instructions Provided to Subjects 15
A.1 Instructions for JNR-ITScam 15
A.2 Instructions for JNR-ITSps 16
REPORT DOCUMENTATION PAGE 17

Figures

Figure 1. Subjective testing was conducted in an office with indirect natural light and sit-stand desks. Credit: Margaret H. Pinson.....9

Tables

Table 1. Scene pools with SRC MOS predictions from design phase. 4
Table 2. HRCs in the JNR-ITSps dataset. 7
Table 3. Gender demographics. 10
Table 4. Age demographics. 10
Table 5. Vision self-assessment demographics 10
Table 6. Number ratings skipped. 10

Preface

This dataset was made possible by the generous assistance of more than several hundred first responders across the U.S. who explained the challenges they encounter when using video systems; by the many first responders, related professionals, and volunteers who agreed to be filmed while simulating typical response scenarios; and by the many product developers, policymakers, researchers, and videographers who shared their perspectives on first responder equipment, video systems, and challenges.

NTIA–ITS foundational work on first responder perspectives on video quality has been supported over the past 30 years by the U.S. Department of Homeland Security (DHS) and the National Institute of Standards and Technology (NIST).

Glossary of Terms

Absolute Category Rating (ACR)	As defined in ITU-T Rec. P.10, *Vocabulary for performance, quality of service and quality of experience* (11/2017): Test method in which subjects are asked to express classification opinions by using absolute quality scales (excellent, good, ...) for their judgment. [1]
Camera Capture Impairments	Any video or image degradation or distortion introduced by camera, at any point between light hitting the lens and creation of the compressed media file. Sources for impairments include the lens, shutter, aperture, viewfinder, focus, image sensor, stabilization, lighting, color filter, image signal processing, resolution, frame rate, digitization, color space, white balance, black balance, color correction, bit depth limitations, chroma subsampling, bandwidth limitations, encoder, and storage.
Controlled Environment	As defined in ITU-T Rec. P.910, *Subjective video quality assessment methods for multimedia applications* (07/2026): A non-distracting environment where a person would reasonably use the device under test (e.g., a lab with controlled lighting). [2]
Hypothetical Reference Circuit (HRC)	As defined in ITU-T Rec. P.910: A fixed combination of a video encoder operating at a given bit rate, network condition and video decoder. The term HRC is preferred when vendor names should not be identified.
Mean Opinion Score (MOS)	As defined in ITU-R Rec. BT.1683, *Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference* (06/2004): The average subjective quality judgment assigned by a panel of viewers (or listeners) to a processed video (or respectively audio) clip. [3]
Processed Video Sequence (PVS)	As defined in ITU-T Rec. P.910: The impaired version of a video sequence.
Reference	As defined in ITU-T Rec. P.910: The original version of each source stimulus. This is the highest quality version available of the audio sample, video clip or audiovisual sequence.

Shot	As defined in ITU-T Rec. P.910: A data source that is a continuous portion of a sequence or stimulus which contains no cuts (e.g., recorded continuously from a live stream).
Source (SRC)	As defined in ITU-T Rec. P.910: A sequence that constitutes the input to a processing system (hypothetical reference circuit) that creates other sequences. NOTE – A source can be used to generate multiple new sequences, from which new stimuli (PVSs) are generated. Example sources include raw footage from a camera, a movie, and computer generated content.
Stimulus	As defined in ITU-T Rec. P.910: Audio sequence, video sequence or audiovisual sequence that is shown to subjects.
Unrepeated Scene Experiment Design	As defined in ITU-T Rec. P.910: A subjective media quality test where each subject views each source stimulus only once.
User-Generated Content (UGC)	Images and videos created by consumers (i.e., not created by professional photographers or videographers).
UYVY Format	UYVY packs 4:2:2 chroma subsampled video pixels into four bytes in the order U, Y, V, Y. Chroma (U and V) are sampled at the location of the first Y pixel.

Executive Summary

From 2025 to 2026, ITU-T Study Group 12 Question 19 performed independent validation testing of models that assess video quality using only pixel values, referred to as "no-reference" or NR models. This effort culminated in ITU-T Rec. J.344, *No-reference objective video quality assessment models for full high-definition video* (07/2026), In force (prepublished) [5].

This report describes a video quality dataset that NTIA–ITS created in support of the ITU-T Study Group 12 Question 19 validation testing of no-reference (NR) models. The dataset name, JNR-ITSps, is a portmanteau (blend) of the ITU project (J.noref), the organizational attribution (NTIA–ITS), and the focus of the dataset (the public safety use case).

JNR-ITSps examines interactions between modern cameras and video systems for the public safety use case. The dataset contains a mixture of images, pristine videos, camera capture impairments, camera coding artifacts, and transcoding artifacts. The JNR-ITSps videos were edited from content that ITS filmed over the past decade to better understand the video quality needs of first responders. The videos represent camera systems and subject matter typically used for first responder tasks.

The videos were rated with the Absolute Category Rating (ACR) method from ITU-T Rec. P.910. The variance in subject ratings is higher than coding-only datasets, which is a known trend in datasets that contain camera capture impairments.

JNR-ITSps was designed to support the training and testing of NR models on contemporary camera systems. The matrix design helps researchers identify model weaknesses to specific cameras and scenarios. The dataset is not intended for comparing the performance of specific cameras, which would require tighter control on filming.

The JNR-ITSps dataset is freely available on the Consumer Digital Video Library (CDVL) (www.cdvl.org). ITU-T Rec. J.344 refers to this dataset as Track 2-2.

JNR-ITSps: Dataset Track 2-2 from ITU-T Study Group 12 Validation of No-Reference Metrics for Camera Capture Impairments, Depicting the Public Safety User Case

Margaret H. Pinson, John Dumke[1], and Rob Grosso[2]

Abstract: This memorandum documents a video quality subjective test conducted by NTIA–ITS in support of the ITU-T Study Group 12 Question 19 validation of no-reference (NR) metrics from 2025 to 2026. This dataset focuses on the public safety use case and is named JNR-ITSps. ITU-T Rec. J.344 refers to this dataset as Track 2-2. The videos demonstrate a mixture of camera capture and transcoding impairments.

Keywords: ACR, camera, ITU-T Study Group 12, no-reference (NR) metric, public safety, subjective test, validation, video quality

[1] The authors are with the U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, Boulder, CO 80305.
[2] The author is formerly with the U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, Boulder, CO 80305.

1. Introduction

This report describes a video quality subjective test that NTIA–ITS conducted in support of the ITU-T Study Group 12 Question 19 validation testing of no-reference (NR) metrics from 2025 to 2026. This effort, referred to as J.noref, is the first independent validation test that was designed specifically for NR metrics. NTIA–ITS conducted two experiments in support of this model validation effort. The first dataset, JNR-ITScam, is described in [4]. The second dataset, JNR-ITSps, is described in this report.

The JNR-ITSps dataset contains a mixture of images, pristine videos, camera capture impairments, camera coding artifacts, and transcoding artifacts. The videos represent camera systems and subject matter typically used for first responder tasks (e.g., video surveillance, drone footage, and dashcams). The JNR-ITSps dataset builds on prior ITS research into the video quality needs of first responders [6]–[16]. Major impairments include camera jiggle, too-fast camera pans, rescaling, and video transcoding artifacts. The dataset includes H.264, H.265, H.266, AV1, and VP9 coding from four different vendors' software encoders.

The JNR-ITSps videos were collected using convenience sampling (i.e., content already owned), which gives the dataset a somewhat eclectic appearance. The original footage was collected over the past decade on professional camcorders, 4K drones, high-performance action cameras, and a few mid-range consumer cameras. Much of this content was 720p and could not be used, because the J.noref test plan mandated a minimum resolution of 1080p. This eliminated all camera recordings that differed substantially from contemporary cameras.

Most of the video content in JNR-ITSps was not used for any prior subjective tests or metric development efforts. About 10% of footage in JNR-ITSps appears in our prior subjective tests [11]–[13][3]. This overlap was considered acceptable for the following reasons. First, the content was edited anew from the raw footage, so that stimuli may overlap but are not identical. Second, the transcoding impairments are entirely different. Third, the videos depict valuable scenarios and impairments that would otherwise not appear in JNR-ITSps, which consumed all of the unforeseen public safety footage that NTIA–ITS had held in reserve.

More information on J.noref can be found in ITU-T Rec. J.344, *No-reference objective video quality assessment models for full high-definition video* (07/2026), In force (prepublished) [5].

[3] We ignore the overlap with [15] for this estimation, because [15] reports full reference metric values without subjective ratings.

2. SRC Videos

JNR-ITSps uses an unrepeated source experiment design. The dataset contains 400 unique SRCs, each depicting simulated public safety content. This content was filmed by 12 different videographers on at least 17 high-performance cameras (e.g., drones, bodycams, handheld cameras, phones, and professional cameras). Some of the videos were taken from [11]–[20]. Other videos were commissioned by ITS over the past decade while exploring the problems that first responders have with video systems. Information on the make and model of the various cameras is not available. As specified in the J.noref test plan, the original content was limited to resolution of full HD 1920×1080 or higher and frame rates of 23.98 fps or higher. The videos depict simulated first responder activities, first responder training events, and live events filmed in public areas (e.g., a wildfire filmed from a highway).

These first responder videos were converted, using FFmpeg [21], to create 400 unencoded SRC videos in the UYVY color data format. These videos are uncompressed and lost no quality in this conversion. Each of these 400 videos is approximately 5-seconds in duration. Small deviations in duration are allowed. While selecting and editing these SRCs, one researcher assigned each SRC a rating on the 5-point scale used by the ACR method, in which excellent=5, good=4, fair=3, poor=2, and bad=1. These ratings helped us estimate the expected distribution of mean opinion scores (MOSs). We will refer to these ad-hoc ratings as $\widehat{ACR}$ to differentiate them from the actual subjects' ratings. In retrospect, these design-phase quality estimations are much lower than the actual MOSs.[4]

The 400 SRCs can be divided into three broad categories:

- Category #1: User-generated content (UGC) with camera capture impairments and coding artifacts from the original camera recording.
- Category #2: UGC with camera capture impairments.
- Category #3: Broadcast-quality content (i.e., a broadcast-quality cameras operated by professional videographers).

Category #1 SRCs span the full range from bad to excellent ($\widehat{ACR} \approx$ 1 to 5). Category #1 SRCs depict the confounding effects of camera capture and coding impairments in contemporary camera systems. Category #1 includes still images.

Category #2 SRCs range from good to excellent ($\widehat{ACR} \approx$ 3 to 5). Most of these videos were filmed by drones and bodycams. Coding artifacts are rarely visible. Category #2 SRCs may contain camera jiggle, too-fast pans, jerky camera movements, white level balance problems, water droplets on the camera lens, and light/shadow interference patterns caused by the drone blades.

[4] People typically agree on quality relationships (e.g., A is better than B) but may disagree on absolute quality ratings (e.g., A is fair quality). The latter question is influenced by many factors, such as consumer expectations, the quality range of stimuli in a subjective test, use case, and an individual's personal biases.

Category #3 SRCs are typically excellent quality ($\widehat{ACR} \approx 5$). Category #3 SRCs contain occasional camera capture artifacts that we expect from broadcast cameras (e.g., moiré interference patterns).

Some of the category #2 and #3 SRCs need transcoding to produce a full range of quality from bad to excellent. The SRCs that needed transcoding were divided into scene pools A through M, where each scene pool depicts similar types of content (see Table 1). The column # SRC lists the number of SRCs in the scene pool. The column $\approx\widehat{MOS}$ averages $\widehat{ACR}$ across that scene pool (i.e., estimated quality *before* transcoding). The $\widehat{MOS}$ and $\widehat{ACR}$ have no residual value and are provided only to explain our process in creating the JNR-ITSps dataset.

The column Transcode identifies whether or not transcoding was applied to the scene pool. A total of 173 videos remain in their original format, with no transcoding; these do not appear in the scene pools. This includes all of the Category #1 SRCs (which evenly span the full range of quality). Some SRCs from categories #2 and #3 also remain in their original format, to establish the camera capture quality baseline.

Table 1. Scene pools with SRC $\widehat{MOS}$ predictions from design phase.

Cat	Scene Pool	# SRC	$\approx\widehat{MOS}$	Transcode
A	City from drone	6	5.00	Yes
B	Car fire response training	8	4.75	Yes
C	Plants and water with ripples	8	4.88	Yes
D	Search and rescue, drone, less movement	12	4.92	Yes
E	Search and rescue, drone, more movement	36	3.53	Yes
F	Search and rescue, bodycam	19	3.74	Yes
G	Moving car, bodycam	9	3.33	Yes
H	Firefighter training, drone and tripod	21	4.71	Yes
I	Firefighter training, burning vehicle	32	4.75	Yes
J	Firefighter training, moiré pattern	23	4.78	Yes
K	Fireground, UGC with shaky camera	27	3.19	Yes
L	Falling rain or snow with flashing lights	12	4.17	Yes
M	Crime scenes	14	4.43	Yes
—	*None*	173	3.21	No

3. Transcoding Impairments

Our goal with transcoding was to select combinations of codecs, bit rates, and resolutions that will be popularly used by first responders in the near future (5 to 10 years). The bulk of first responder departments' funds must be devoted to high price items such as vehicles and salaries [8]. Therefore, the first responder community relies primarily on video systems developed for other commercial applications, and thus first responder videos must be considered a subtype of UGC. The JNR-ITSps dataset focuses on contemporary camera systems and avoids outdated impairments (e.g., obvious blocking artifacts).

In early November 2025, we performed a search of common video streaming and video hosting services, to find what combinations of bit rate, codec, and resolutions were commonly used. Of particular interest to us were devices used by public safety (e.g., bodycam, in-car cameras, surveillance systems, video storage for retention of first responder videos). Bodycams have improved rapidly over the past decade. While older bodycams used lower resolutions, the minimum standard is now 1080p 30 fps with several hours of recording capacity. Consumer content streaming services are of interest to first responders, because some departments in California use these services to distribute first responder videos to the public. These product specifications informed our experiment design.

Table 2 lists the final selection of hypothetical reference circuits (HRC). The column Count contains the number of PVSs associated with this HRC. The column Software identifies the software used for transcoding. Our prior experience indicates that differences among software will produce different encoding artifacts. The J.noref test plan placed constraints on the codecs, bit rates, and resolutions.

HRC 1 contains coding and camera capture impairments only. The editing bit rate was chosen to be high enough to avoid transcoding impairments. HRC 1 is not, strictly speaking, one system, because it contains videos from various cameras. As noted above, HRC 1 includes SRCs from Category 1 and SRCs selected from scene pools A through M. Other SRCs were assigned to HRC 1 because the SRCs contained unusual and interesting shapes, movements, or camera capture impairments (e.g., smoke obscuring objects of interest, objects that move past each other, dirty glass between the camera and objects of interest, visual impairments caused by a drone's flight mechanism, moiré interference pattern, and white level balance problems).

HRCs 2 to 11 use H.264, H.265, and AV1 transcoding from FFmpeg. These HRCs were influenced by the VCRDCI dataset[5] which shows a "knee" immediately below 27 constant rate factor (CRF) for H.264, H.265, and AV1. From 18 CRF to 27 CRF, the quality drops gradually. Below 27, CRF quality drops rapidly. FFmpeg was chosen because:

[5] Robert Grosso and Margaret H. Pinson, "VMAF Compression Ratings that Disregard Camera Impairments (VCRDCI) Dataset," Technical Memorandum NTIA TM-22-564, U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, September 2022. https://doi.org/10.70220/kqqzgrqv

- FFmpeg allows the user to control encoding using quantization parameter (QP) levels, which produces similar quality across an HRC (but diverse bit rates).
- FFmpeg is free, light-weight, highly compatible, and popular.
- Many free video converters use FFmpeg.

The HRCs 13 to 20 use H.264 and H.265 transcoding from Adobe Premiere Pro. They also use constant bit rate encoding (CBR), which is expected to produce diverse quality and similar bit rates among SRCs. The codec H.264 appears more frequently than H.265 because H.264 is more widely used by camera systems. The codec H.265 was encoded at 1920×1080 resolution for all bit rates on the advice of other J.noref proponents[6], who noted that H.265 effectively adjusts the video resolution during the encoding process. Adobe Premiere Pro was chosen because:

- Adobe Premiere Pro has existed for more than 15 years.
- Adobe Premiere Pro is popular with professional content creators.
- Adobe Premiere Pro is an expensive encoder and video editing system that supports H.264 and H.265.

Video coding standards typically only specify the decoder. The encoder implementation is chosen by the vendor, offering opportunity for competitive advantage. Thus, the 15 year history and higher price suggest a well-designed set of coding algorithms.

HRCs 21 to 28 use AV1 and VP9 transcoding from TMPGEnc Video Mastering Works 8. The bit rates are intentionally lower than H.264 bit rates associated with HRCs 13 to 18, because AV1 and VP9 are newer codecs. TMPGEnc was chosen because:

- TMPGEnc has existed for more than 15 years.
- TMPGEnc is designed to encode multiple video sequences with diverse encoding settings.
- TMPGEnc is an inexpensive encoder that supports AV1 and VP9.

HRCs 29 to 34 use VVC transcoding from MainConcept, which was the only commercial vendor we could find that supported VVC encoding. Currently, MainConcept only offers VVC for live streaming. MainConcept supported the development of the JNR-ITSps dataset dataset by privately creating the HRCs using its own best practices to choose bit rates, resolutions, filters, and other encoder settings. All VVC streams were encoded using MainConcept's Variable Bit Rate (VBR) rate control mode. MainConcept converted all of the video streams to 25 fps[7]. MainConcept was chosen because:

- MainConcept has existed for more than 15 years.

[6] ITU-T Study Group 12 uses the term "proponents" to refer to organizations that submit a model for validation.
[7] Europe uses 50 Hz power, so video frame rates are typically multiple of 25. The US uses 60 Hz power, so video frame rates are typically multiple of 30 (which are vulnerable to flickering, banding, and strobing artifacts) or 29.97 (to avoid these interference and light flicker impairments). Cinema typically use 24 fps.

- MainConcept software (not used) is designed to encode multiple video sequences with diverse encoding settings.
- MainConcept offered subject matter experts who could identify and select a robust variety of transcoding settings. This improved value of the dataset for independent validation.

The JNR-ITSps dataset contains video sequences with 23.98, 24, 25, 29.97, and 59.94 fps.

Table 2. HRCs in the JNR-ITSps dataset.

HRC	Software	Codec	Resolution	Encode	Count
1	NA	Camera	1920×1080	NA	184
2	FFmpeg	H.264	1280×720	27 CRF	7
3	FFmpeg	H.264	854×480	27 CRF	7
4	FFmpeg	H.264	640×360	27 CRF	7
5	FFmpeg	H.265	1920×1080	27 CRF	8
6	FFmpeg	H.265	1280×720	27 CRF	7
7	FFmpeg	H.265	854×480	27 CRF	7
8	FFmpeg	H.265	640×360	27 CRF	7
9	FFmpeg	AV1	1920×1080	27 CRF	7
10	FFmpeg	AV1	1280×720	27 CRF	7
11	FFmpeg	AV1	854×480	27 CRF	7
12	FFmpeg	AV1	640×360	27 CRF	7
13	Adobe Premiere Pro	H.264	1920×1080	8 Mbps CBR	7
14	Adobe Premiere Pro	H.264	1920×1080	5 Mbps CBR	7
15	Adobe Premiere Pro	H.264	1280×720	5 Mbps CBR	7
16	Adobe Premiere Pro	H.264	1280×720	3 Mbps CBR	7
17	Adobe Premiere Pro	H.264	854×480	4 Mbps CBR	7
18	Adobe Premiere Pro	H.264	854×480	2 Mbps CBR	7
19	Adobe Premiere Pro	H.265	1920×1080	2 Mbps CBR	7
20	Adobe Premiere Pro	H.265	1920×1080	1 Mbps CBR	8
21	TMPGEnc	AV1	1920×1080	5 Mbps CBR	8
22	TMPGEnc	AV1	1280×720	3 Mbps CBR	7
23	TMPGEnc	AV1	854×480	2 Mbps CBR	7
24	TMPGEnc	AV1	640×360	1 Mbps CBR	7
25	TMPGEnc	VP9	1920×1080	5 Mbps CBR	7
26	TMPGEnc	VP9	1280×720	3 Mbps CBR	8
27	TMPGEnc	VP9	854×480	2 Mbps CBR	7
28	TMPGEnc	VP9	640×360	1 Mbps CBR	8
29	MainConcept	VVC	1920×1080 25 fps	4.5 Mbps VBR	4
30	MainConcept	VVC	1920×1080 25 fps	3 Mbps VBR	4

HRC	Software	Codec	Resolution	Encode	Count
31	MainConcept	VVC	1280×720 25 fps	2.2 Mbps VBR	4
32	MainConcept	VVC	1280×720 25 fps	1.6 Mbps VBR	3
33	MainConcept	VVC	960×540 25 fps	1 Mbps VBR	4
34	MainConcept	VVC	640×360 25 fps	0.1 Mbps VBR	3

The HRCs that required transcoding were randomly assigned to scene pools A through M with the constraints that (1) the null impairment must be selected exactly once for each scene pool, and (2) the same HRC cannot be assigned to two consecutively numbered SRCs. The null impairment (i.e., the original camera recording) is included to ensure that the highest quality level is included for each scene pool. The numbering constraint was imposed because most SRCs were selected from longer shots (filmed continuously), so SRCs with consecutive numbers are likely to have very similar content. To ensure a similar number of PVSs were assigned to each HRC, the list of HRCs was randomly permuted and assigned in that order to the scene pools.

After this initial assignment, $\widehat{MOS}$ was calculated for the HRC's SRC, by taking the mean of the SRCs $\widehat{ACR}$ described in Section 2. If any HRC's $\widehat{MOS}$ fell below 4.0, we found a SRC in that scene pool with a low $\widehat{ACR}$ (1, 2, or 3) and switched its HRC assignment with a SRC that had a high $\widehat{ACR}$ (4 or 5). This swapping process continued until each scene pool had an estimated $\widehat{MOS}$ of 4.0 or higher. The goal was to avoid any HRC being associated with too many SRCs that contain fast pans, spins, camera capture, or other quality problems associated with the original recording. By the end of this pseudo-random HRC assignment process, 7 or 8 SRCs were associated with each HRC.

MainConcept provided more HRCs than we expected (six instead of three). Therefore, HRCs 29, 30, and 31 were each split into two HRCs and a similar swapping was performed to ensure that each HRC had an estimated $\widehat{MOS}$ of 4.0 or higher. These six HRC pools were renumbered HRCs 29 to 34 (with 3 or 4 HRCs each).

Unfortunately, the J.noref official data analysis eliminated the VVC HRCs due to the low number of subject ratings and retained only the 24 subjects with the highest Pearson correlation to mean opinion score (MOS) computed from all subjects.

4. Subjective Testing

Appendix A contains instructions provided to the test subjects. Readers who prefer to learn by example may want to read those instructions before this section.

The subjective test was conducted on a 27-inch Dell monitor with1920×1080 resolution. Figure 1 shows the quiet room used for subjective testing. This room contained the standard office configuration for our office: a quiet room with sit-stand desks, ergonomic chairs, fluorescent lights, and indirect natural light coming through large windows. The temp agency provided 28 subjects with a broad distribution of genders and ages (see Table 3 and Table 4). The data spreadsheet also contains ratings from one pre-test subject (29) who used a 13-inch laptop monitor in a different location and is omitted from the demographics reported below.

Figure 1. Subjective testing was conducted in an office with indirect natural light and sit-stand desks. Credit: Margaret H. Pinson

The experiment was controlled by the AVRateNG software [22]–[23]. The experiment used the Absolute Category Rating (ACR) method with a 5-point scale and the skip option.

The experiment was self-paced. Subjects were paid for four hours, encouraged to take breaks, and left as soon as they finished. Most subjects took between 2 and 2.5 hours to finish both experiments, and very few subjects took longer than 3 hours.

Due to a problem with file conversions, the playback of HRCs 29 to 34 were corrupted during most of the subjective testing. This issue was discovered about two-thirds of the way through the experiment. Consequently, only 10 ratings are recorded for these six VVC HRCs.

Table 3 shows the distribution of subject genders, Table 4 shows the distribution of subject ages, and Table 5 shows the distribution of subject visual acuity (self-assessment). Table 6 shows the distribution of number of PVSs skipped. This shows that the skip option was not abused: 57% of subjects never used the skip option and the maximum was 4% of PVSs skipped.

Table 3. Gender demographics.

Male	Female	Non-binary	Rather not say
15	11	1	1

Table 4. Age demographics.

18 to 24	25 to 29	30 to 39	40 to 49	50 to 59	60 to 69
9	3	9	4	2	1

Table 5. Vision self-assessment demographics.

Excellent	Good	Fair	Poor	Bad
10	14	4	0	0

Table 6. Number ratings skipped.

0 Skip	1 Skip	2 Skip	3 Skip	4 Skip	5 Skip	16 Skip
16	4	1	3	1	2	1

5. Conclusions

The JNR-ITSps dataset is available in the Consumer Digital Video Library (CDVL, www.cdvl.org). The ZIP file contains the following:

- 400 videos rated by subjects
- Spreadsheet with individual subject ratings
- Spreadsheet identifying the PVSs and subjects that were used for the J.noref analyses

ITU-T Rec. J.344 refers to JNR-ITSps as Track 2-2.

6. References

[1] International Telecommunication Union, Recommendation ITU-T P.10 (11/2017), *Vocabulary for performance, quality of service and quality of experience*, ITU, Geneva, Switzerland. https://www.itu.int/rec/T-REC-P.10

[2] International Telecommunication Union, Recommendation ITU-T P.910 (07/2026), *Subjective video quality assessment methods for multimedia applications*, ITU, Geneva, Switzerland. https://www.itu.int/rec/T-REC-P.910

[3] International Telecommunication Union, Recommendation ITU-R BT.1683 (06/2004), *Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference*, ITU, Geneva, Switzerland. https://www.itu.int/rec/R-REC-BT.1683/en

[4] Margaret H. Pinson and Rob Grosso, "JNR-ITScam: Dataset Track 2-1 from ITU Study Group 12 Validation of No-Reference Metric for Camera Capture Impairments, Depicting the User Generated Content Use Case," Technical Memorandum NTIA TM-27-584, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, October 2026. https://doi.org/10.70220/nxkzt99s

[5] International Telecommunication Union, Recommendation ITU-T J.344 (07/2026), In force (prepublished), *No-reference objective video quality assessment models for full high-definition video*, ITU, Geneva, Switzerland. https://www.itu.int/rec/T-REC-J.344-202607-P/en

[6] Margaret H. Pinson, Stephen Wolf, and Robert B. Stafford, "Video Performance Requirements for Tactical Video Applications," in 2007 IEEE Conference on Technologies for Homeland Security (Woburn, Massachusetts, May 2007), 85–90. https://doi.org/10.1109/THS.2007.370025

[7] Joel Dumke, "Visual acuity and task-based video quality in public safety applications," in *Proc. of SPIE-Image Quality and System Performance X, Electronic Imaging* SPIE Vol. 8653, 865306, Burlingame, CA, 5–7 February 2013. https://doi.org/10.1117/12.2004882

[8] Margaret H. Pinson, "Technology Gaps in First Responder Cameras," Technical Memorandum NTIA TM-17-524, U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, May 2017. https://doi.org/10.70220/l83fl1hk

[9] Margaret H. Pinson, "Gaps in Public-Safety Cameras: How Dust, Rain, Snow, and Shadows Impact Public Safety Video." *MissionCritical Communications* (August 2017): 30–35. https://its.ntia.gov/publications/3192.aspx

[10] Margaret H. Pinson, "Caught in the Act: Law-Enforcement Feedback on Video Surveillance Recordings," in *IEEE Consumer Electronics Magazine* 8, no. 1 (January 2019): 82–86. https://doi.org/10.1109/MCE.2018.2867981

[11] Margaret H. Pinson, "ITS4S2: An Image Quality Dataset With Unrepeated Images From Consumer Cameras," Technical Memorandum NTIA TM-19-537, U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, April 2019. https://doi.org/10.70220/478jr91r

[12] Margaret H. Pinson, "ITS4S3: A Video Quality Dataset With Unrepeated Videos, Camera Impairments, and Public Safety Scenarios," Technical Memorandum NTIA TM-19-538, U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, April 2019. https://doi.org/10.70220/3rlnl9jw

[13] Margaret H. Pinson and Samuel Elting, "ITS4S4: A Video Quality Study of Camera Pans," Technical Memorandum NTIA TM-20-545, U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, December 2019. https://doi.org/10.70220/740j2cvp

[14] Margaret H. Pinson, "Analysis of No-Reference Metrics for Image and Video Quality of Consumer Applications," Technical Memorandum NTIA TM-20-547, U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, January 2020. https://doi.org/10.70220/hb2drbf6

[15] Robert Grosso and Margaret H. Pinson, "VMAF Compression Ratings that Disregard Camera Impairments (VCRDCI) Dataset," Technical Memorandum NTIA TM-22-564, U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, September 2022. https://doi.org/10.70220/kqqzgrqv

[16] Margaret H. Pinson, "Why No Reference Metrics for Image and Video Quality Lack Accuracy and Reproducibility," *IEEE Transactions on Broadcasting* 69, no. 1 (March 2023): 97–117. https://doi.org/10.1109/TBC.2022.3191059

[17] Marius Facktor, Abhinau K. Venkataramanan, Praful Gupta, and Alan C. Bovik, "RoadImpairs: A Roadway Image Dataset with Impairments and Object-Detection Failure Rate Estimation," The Consumer Digital Video Library (access limited to account holders), May 5,2021. https://www.cdvl.org/members-section/view-file/?videoid=3033

[18] Hang Shi and Chengjun Liu, "An Innovative Video Quality Assessment Method and An Impairment Video Dataset", 2021 IEEE International Conference on Imaging Systems and Techniques (IST) (Kaohsiung, Taiwan, August 2021), 1–6. https://doi.org/10.1109/IST50367.2021.9651418

[19] Jeffrey Liu, David Strohschein, Siddharth Samsi, and Andrew Weinert "Large Scale Organization and Inference of an Imagery Dataset for Public Safety," IEEE High Performance Extreme Computing Conference (HPEC) (Waltham, Massachusetts, September 2019), 1–6. https://doi.org/10.1109/HPEC.2019.8916437

[20] Andrew Weinert, "Video and Imagery Dataset to Drive Public Safety Capabilities," Public Safety Broadband Stakeholder Meeting, July 10, 2019. Accessed October 15,

2025: Event description https://www.nist.gov/ctl/pscr/psiap-new-jersey and slide deck https://www.nist.gov/system/files/documents/2019/11/01/video_and_imagery_dataset_to_drive_public_safety_capabilities.pdf

[21] FFmpeg Developers, "FFmpeg," Software, *FFmpeg.org*. https://ffmpeg.org/.

[22] Steve Göring, Rakesh Rao Ramachandra Rao, Stephan Fremerey, and Alexander Raake, "AVrate Voyager: An Open Source Online Testing Platform," in *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)* (Tampere, Finland), 1–6. https://doi.org/10.1109/MMSP53017.2021.9733561

[23] "AVRateNG – GitHub Project," https://github.com/Telecommunication-Telemedia-Assessment/avrateNG. Accessed: January 5, 2026.

[24] Margaret H. Pinson et al., "The Influence of Subjects and Environment on Audiovisual Subjective Tests: An International Study," *IEEE Journal of Selected Topics in Signal Processing* 6, no. 6 (October 2012): 640–651. https://doi.org/10.1109/JSTSP.2012.2215306

Appendix A. Instructions Provided to Subjects

This Appendix contains the instructions for two experiments: JNR-ITScam [4] and JNR-ITSps (described in this memorandum). Subjects participated in JNR-ITScam and then JNR-ITSps.

We conducted JNR-ITScam first, because it has a narrower range of quality than JNR-ITSps. Subjects always adapt the rating scale to the context of videos presented. MOSs are relative, not absolute [24]. By conducting JNR-ITScam first, we allowed for as wide a range of MOSs as possible for JNR-ITScam.

A.1 Instructions for JNR-ITScam

Thank you for participating in our study. The purpose of this study is to gather individual perceptions of the quality of several short videos. Your data will support an international effort to validate the accuracy of metrics that estimate video quality.

In this experiment, you will be presented with a series of 5 second videos. Each time a video is played, you will be asked to judge its quality. A rating screen will appear with the options of excellent, good, fair, poor, and bad. Use the mouse to select the rating that best describes your opinion of the video. After you have clicked on one of the options, click on the "Rate" button to record your response and play the next video.

You can pause the test at any time by entering your quality judgement but not pressing the "Rate" button.

The rating screen includes a "skip" option, in case you get distracted and cannot rate the video. This happens to everyone occasionally.

Do not worry about somehow giving the wrong answer; there is no right or wrong answer. Everyone's opinion will be slightly different. We simply want to record your opinion.

Two warnings about the test software. First, the very first video played and the first rating screen are sometimes not well synchronized. This should not happen again. Second, you can move the rating window, but please don't. We want you to watch the entire 5 second video before rating it.

Today's experiment is divided into two topics. In the first part, the videos show the same scenarios filmed on a variety of modern cameras. One of these videos examines the camera's ability to change focus from a near object (flowers) to a far object (trees). In the second part, the videos show modern video systems used by first responders. I will need to change some settings on your computer before you can proceed to the second part.

We will start with a few practice clips while I am standing here. After that, the experiment will be self-paced. We encourage you take breaks occasionally, but this is not required.

<<play the practice session>>

I will be in the room next door. Come get me if you have any questions, problems, or need to use the restroom. Feel free to adjust the desk, chair, and monitor.

A.2 Instructions for JNR-ITSps

The second part of our test explores the first responder use case: in-car cameras, automated driving, video surveillance, drones, search and rescue, firefighters, law enforcement, training events, etc.

10% of the videos in this session are pictures, because law enforcement officers often need to extract single frames from videos to understand what is going on.

The test interface is the same, but you will see a mixture of camera capture impairments and video compression impairments. As you rate these videos, pretend you are a first responder, and you have to use the video or picture to do your job.

We will start with another short training session, to help you adjust to this scenario and the new range of quality.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
October 2026	Technical Memorandum	01/2025	09/2026

4. TITLE AND SUBTITLE
JNR-ITSps: Dataset Track 2-2 from ITU-T Study Group 12 Validation of No-Reference Metrics for Camera Capture Impairments, Depicting the Public Safety User Case

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
3147012-300	1	NTIA-ITS.TAM

6. AUTHOR(S)
Margaret H. Pinson, John Dumke, and Rob Grosso

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Telecommunication Sciences
National Telecommunications and Information Administration
U.S. Department of Commerce
325 Broadway
Boulder, CO 80305

8. PERFORMING ORGANIZATION REPORT NUMBER
NTIA TM-27-585

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Telecommunications and Information Administration
Herbert C. Hoover Building
1401 Constitution Ave., N.W.
Washington, D.C. 20230

10. SPONSOR/MONITOR'S ACRONYM(S)
NTIA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This paper documents a video quality subjective test conducted by NTIA–ITS in support of the ITU-T Study Group 12 Question 19 validation of no-reference (NR) metrics from 2025 to 2026. This dataset focuses on the public safety use case and is named JNR-ITSps. ITU-T Rec. J.344 refers to this dataset as Track 2-2. The videos demonstrate a mixture of camera capture and transcoding impairments.

15. SUBJECT TERMS
ACR, camera, ITU-T Study Group 12, no-reference (NR) metric, public safety, subjective test, validation, video quality

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
Unclassified	Unclassified	Unclassified	UU	28

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Margaret H. Pinson	*720-601-7314*