ZPEG Engine for x264 Technology Report

Jan Ozer/Streaming Learning Center
@janozer

Table of Contents

Overview

The ZPEG Engine is a video pre-processing technology that improves the encoded quality of codecs like H.264, HEVC, VP9 and MPEG-2. The ZPEG Engine was developed by ZPEG, Inc. According to the company's website, "The ZPEG Engine preprocessing is implemented using a human visual model that optimizes each specific asset (video), providing far lower bandwidths without objectionable artifacts (blocking, stair-stepping, ghosting)."

The tests documented in this report focus on how the ZPEG Engine performed as an upgrade to the widely used x264 video encoder codec. The ZPEG Engine is currently offered as a 100%-compatible technology upgrade to x264. The ZPEG Engine is also applicable to other codecs, but only the x264 version was tested.

CoreCodec, Inc. is an x264, LLC partner and contracted with the Streaming Learning Center (SLC) to test the ZPEG Engine and produce this report. Though test procedures were discussed in advance, positive results were not guaranteed.

In essence, the SLC tests were designed to determine whether the ZPEG technology could substantially reduce the data rate of the test clips without producing objectionable artifacts when compared to x264. Side-by-side testing of x264 and x264 with ZPEG was performed to determine what, if any, advantage is produced by the ZPEG Engine.

How We Tested

ZPEG Inc. provided SLC with a version of the x264 codec compiled with the ZPEG Engine. ZPEG Engine preprocessing was accessible via a simple command line switch whereby the command --zpeg encoded with ZPEG Engine enabled, and the command --no-zpeg encoded with the ZPEG Engine disabled. Separately, SLC tested a publicly available x264-only build to confirm the –no-zpeg encodings.

To test the ZPEG Engine, SLC encoded a library of 15 1080p source videos which contains a mix of movie and movie-like real world footage, multiple animated clips, and business-oriented clips like screencam, PowerPoint-source, and talking head footage. This diverse set of content was designed to identify where the ZPEG Engine worked best and where it didn't.

To test the ZPEG Engine, SLC encoded all source clips with and without using the ZPEG Engine. Since the source footage was so varied, SLC encoded using constant rate factor (CRF) encoding with a value of --CRF 23. CRF is a form of per-title encoding that produces a specific target quality level rather than a specific data rate.

In SLC's experience, a value of CRF 23 produces roughly the quality used by Hollywood for movies and TV shows distributed on iTunes. So, the clips without ZPEG duplicated that quality, while the ZPEG technology attempted to reduce the data rate further without creating noticeable artifacts.

SLC performed and summarized these encodings, and then performed subjective comparisons using tools and techniques discussed in more detail below.

Encoding Time and CPU Consumption

HP ZBook Studio G3

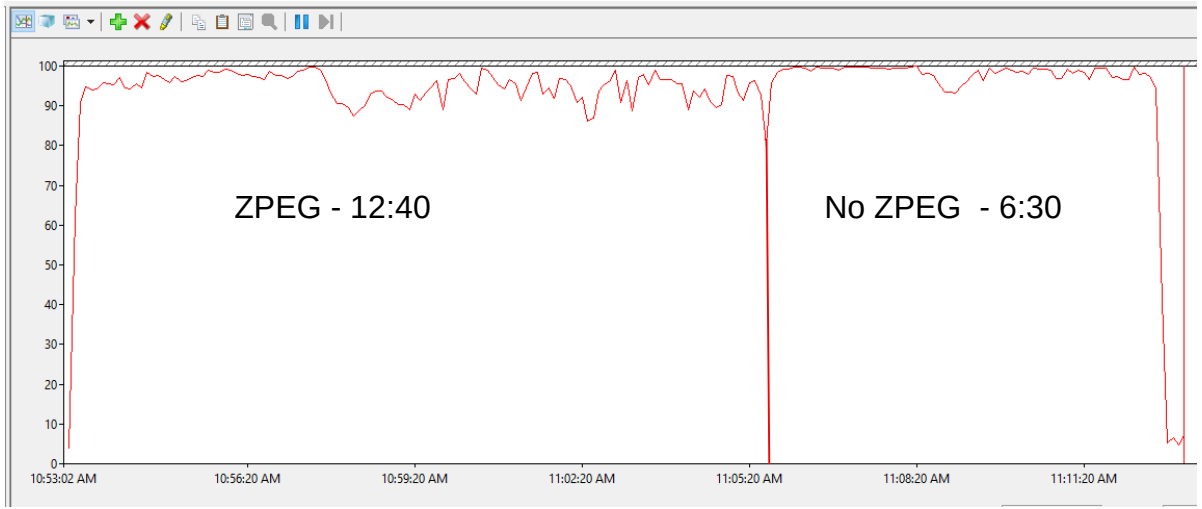
PC name	DESKTOP-0F3TKDV

Rename this PC

Organization	WORKGROUP

Connect to work or school

Edition	Windows 10 Pro
Version	1703
OS Build	15063.909
Product ID	00330-50026-77250-AAOEM
Processor	Intel(R) Xeon(R) CPU E3-1505M v5 @ 2.80GHz 2.81 GHz
Installed RAM	32.0 GB (31.9 GB usable)
System type	64-bit operating system, x64-based processor



We tested on a single CPU HP ZBook Studio G3 notebook running a 2.81 GHz Intel Xeon E3-1505M CPU with 32 GB of RAM.

ZPEG is a relatively CPU intensive process that roughly doubled the encoding time of the test clips. The clip shown above is a five minute segment from the movie Zoolander.

Data Rate Savings

Data rate savings are summarized on the right, with the clips divided into three categories, animated, movie-ish, and business.

ZPEG was most effective with real-world movie-ish clips (27.54% savings over x264), including clips from two actual movies, Elektra and Zoolander.

ZPEG performed well (24.67% on average savings) on business-oriented clips, though some content types like screencams are contra-indicated. Performance on real world clips Epiphan and Talking Head was excellent.

ZPEG processing was least effective with the animated clips (11.52% savings). Even Tears of Steel, which is a mix of CG and real world content, produced only modest benefits

Overall, the average savings over x264 was 21.24%.

	Rez	FPS	No Z	Z	Delta
Movie-ish					
Elektra	1080p	30	4,720	3,290	-30.30%
Freedom	1080p	30	5,877	4,768	-18.87%
Haunted	1080p	30	6,359	4,259	-33.02%
Zoo	1080p	30	8,989	6,476	-27.96%
					-27.54%
Business					
Epiphan	1080p	30	1,934	1,197	-38.11%
New	1080p	30	3,270	2,686	-17.86%
Screencam	1080p	30	704	662	-5.97%
Talking Head	1080p	30	2,860	1,763	-38.36%
Test	1080p	30	4,312	3,004	-30.33%
Tutorial	1080p	30	506	418	-17.39%
					-24.67%
Animated					
El Ultimo	1080p	30	1,903	1,716	-9.83%
Ironman	1080p	30	4,278	4,138	-3.27%
Sintel	1080p	30	5,053	4,539	-10.17%
Sponge Bob	1080p	30	5,459	4,155	-23.89%
Tears of Steel	1080p	30	4,622	4,140	-10.43%
					-11.52%
					-21.24%

Subjective Comparison - No Objectionable Artifacts

The standard applied for these tests was whether the data rate reductions produced by ZPEG were “without objectionable artifacts (blocking, stair-stepping, ghosting).”

In a more general sense, the issue was whether the average viewer watching the video playback in real time would notice that they were watching the ZPEG pre-processed clip rather than the unprocessed clip. In all cases, the answer was no.

Specifically, the data rate reductions that ZPEG achieved were clear of artifacts or other quality deficits that the average viewer would notice. SLC used three types of tests to make this determination.

- Video playback in real time
- Side-by-side analysis in Adobe Premiere Pro; real-time and frame by frame
- Quality-map view in SSIM Plus Quality of Experience Monitor (SQM) where ZPEG produced high data rate reductions

Side-by-Side Displays in Premiere Pro

The side-by-side display in Premiere Pro made it simple to compare the center cut of the two files both frame-by-frame and during realtime playback. For example, the screen on the right (and in full screen on the following slide), from the movie Elektra, shows the ZPEG pre-processed clip on the left.

As you can see, despite a data rate reduction of 30%, the ZPEG-processed clip shows no artifacts of any kind. Color and exposure are also nearly identical.

The only noticeable difference is a very slight reduction of some details, like fewer freckles, and slightly smoother skin. Of course, in the absence of artifacts, viewers don't notice details that aren't there, particularly during real-time playback.

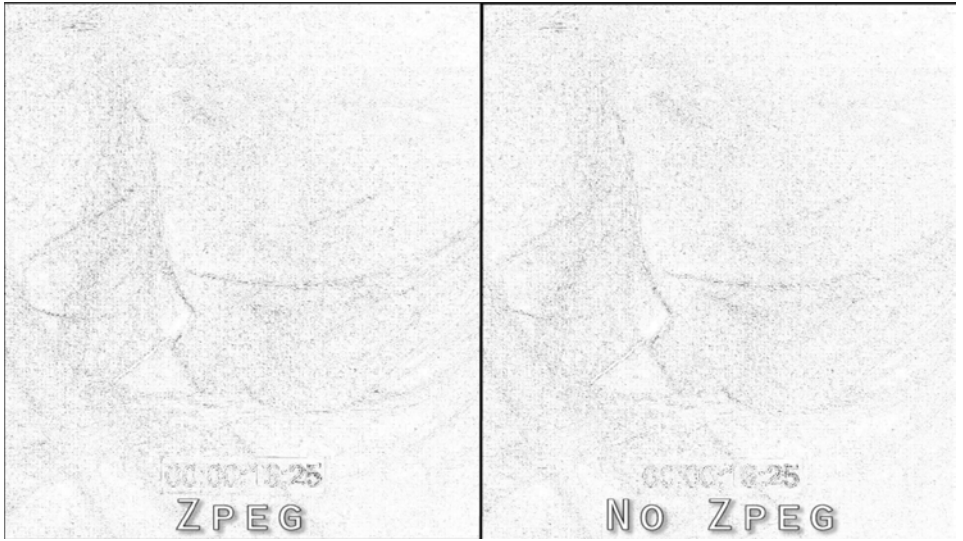
Note that ZPEG can calibrate rocessing to match the viewing distance of the video, so you could process clips viewed on a smartphone differently than those displayed on a TV set. We used the default settings for all clips and performed all visual comparisons on an HP LP3065 monitor from a normal viewing distance.

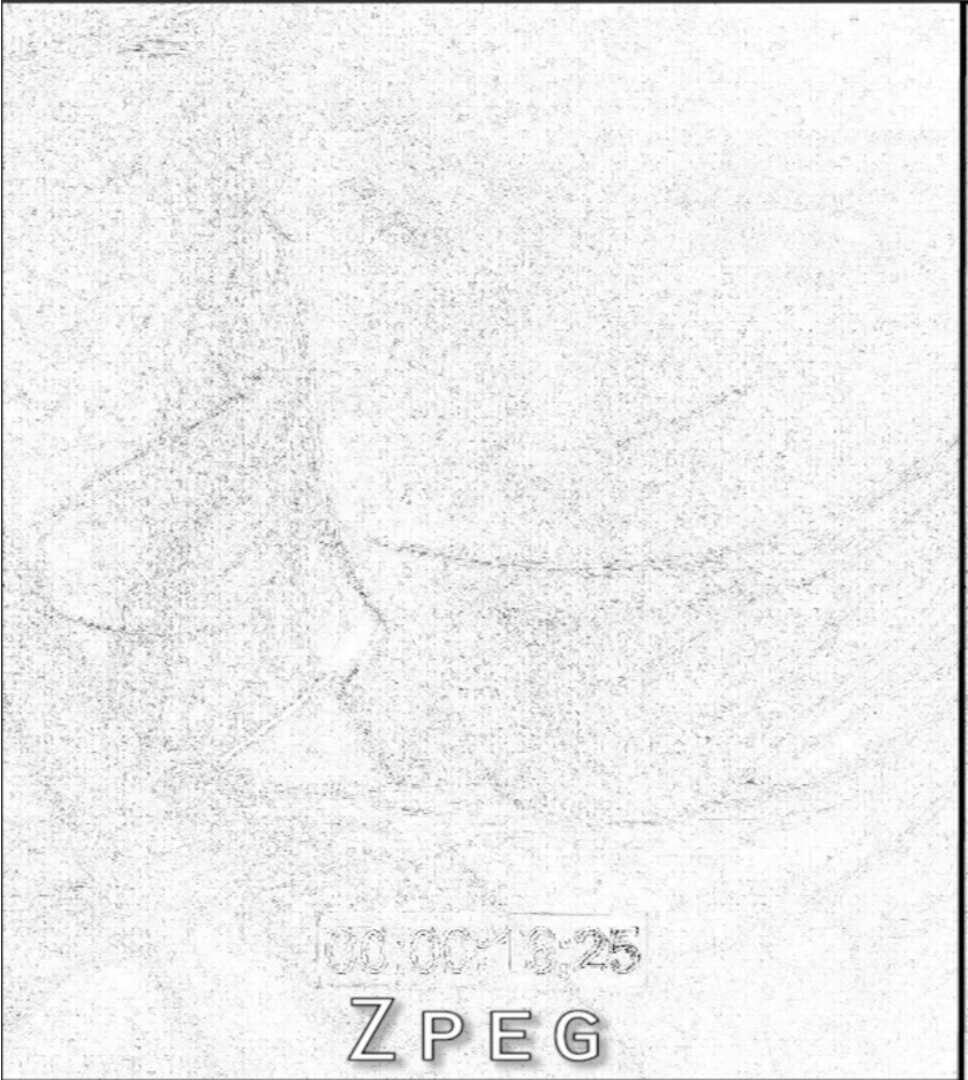

00;00;16;25
ZPEG
00;00;16;25
NO ZPEG

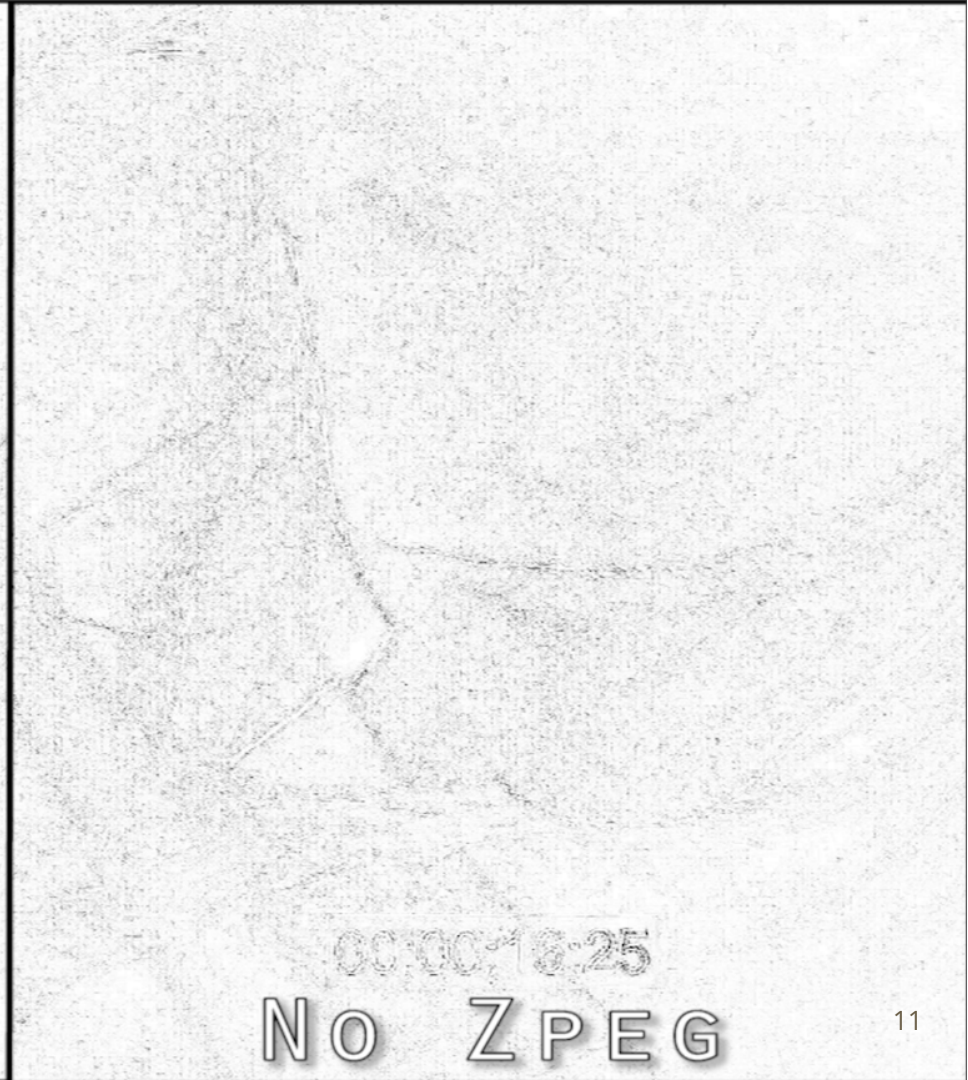
Quality Map View in SQM

The quality map view in the SSIMWave Quality of Experience Monitor (SQM) shows how the processed and unprocessed frames differed from the source. To explain, on the right (and in full screen on the following slide) is an amalgam of two frames; on the left the difference map comparing the ZPEG processed frame and the source; on the right, the difference between the unprocessed frame and the source.

As you can see, the difference maps are virtually identical, confirming the conclusion that the typical viewer watching in real time would see no differences between the processed and unprocessed clips.




00:00:18:25
ZPEG


00:00:18:25
NO ZPEG

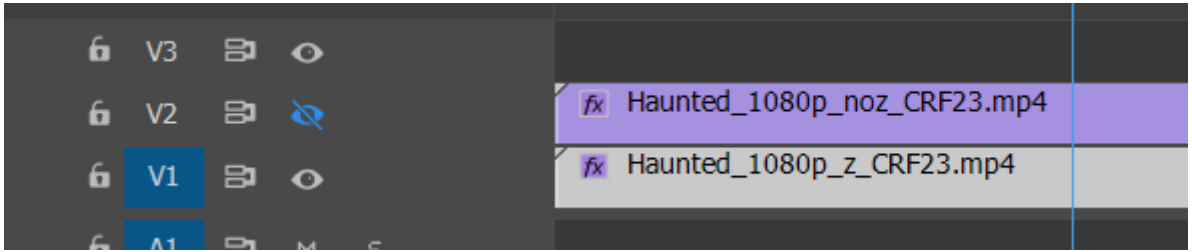
Objective Quality Benchmarks Not Helpful

We measured all clips using objective quality metrics PSNR and VMAF but didn't find the results helpful. That's because when the scores indicated a potential quality deficit in the ZPEG-processed clip, subjective observation didn't confirm these findings. This sometimes occurs when enhancement techniques improve subjective quality but perform operations that trigger lower subjective ratings.

For this reason, most companies marketing optimization technologies recommend against evaluating them with objective benchmarks, and using subjective methods instead.

We used the Moscow State University Video Quality Measurement Tool (above) to identify any regions in the clip where quality might diverge. This compares the PSNR values of the processed (reddish) and unprocessed (green) clips. We checked all frames with significantly downward spikes, like those just before the 1000 frame marker on the left. You'll see such an analysis for each video we tested.

Subjective Comparison - Deeper Analysis



SLC also viewed all files by placing them on a Premiere Pro timeline, one a top the other, and hiding and revealing the top video (top/bottom view). This is the best way to see true differences between the clips and was used for diagnostic purposes.

Viewers watching in real time would not notice any of the differences revealed in this view. Where the differences revealed in this view were significant, SLC recorded narrated Camtasia videos highlighting and discussing them. These videos may be made available upon request to CoreCodec.

Representative Comparisons

At this point, we'll move to a clip by clip analysis. For each clip, we'll have a brief description, and will show representative comparison frames.

For those clips showing higher levels of savings, we'll also show quality maps. We'll review the clips in the order shown on the right.

	Rez	FPS	No Z	Z	Delta
Movie-ish					
Elektra	1080p	30	4,720	3,290	-30.30%
Freedom	1080p	30	5,877	4,768	-18.87%
Haunted	1080p	30	6,359	4,259	-33.02%
Zoo	1080p	30	8,989	6,476	-27.96%
					-27.54%
Business					
Epiphan	1080p	30	1,934	1,197	-38.11%
New	1080p	30	3,270	2,686	-17.86%
Screencam	1080p	30	704	662	-5.97%
Talking Head	1080p	30	2,860	1,763	-38.36%
Test	1080p	30	4,312	3,004	-30.33%
Tutorial	1080p	30	506	418	-17.39%
					-24.67%
Animated					
El Ultimo	1080p	30	1,903	1,716	-9.83%
Ironman	1080p	30	4,278	4,138	-3.27%
Sintel	1080p	30	5,053	4,539	-10.17%
Sponge Bob	1080p	30	5,459	4,155	-23.89%
Tears of Steel	1080p	30	4,622	4,140	-10.43%
					-11.52%
					-21.24%

Elektra

Description: Elektra is a real-world movie starring Jennifer Garner that proves, more than anything, that Ben Affleck is the dumbest guy alive.

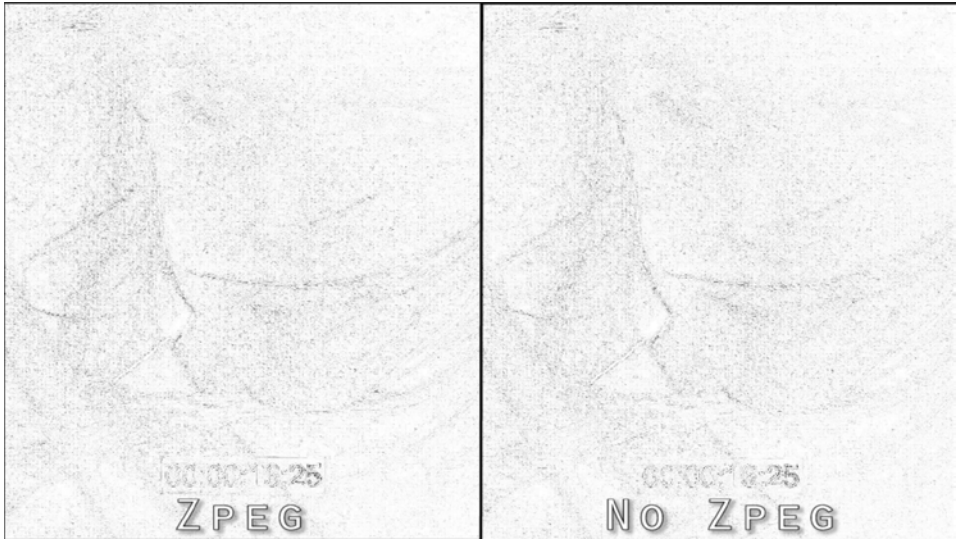
Data rate reduction: -30.30%
VMAF differential: -3.64
PSNR differential: -1.54

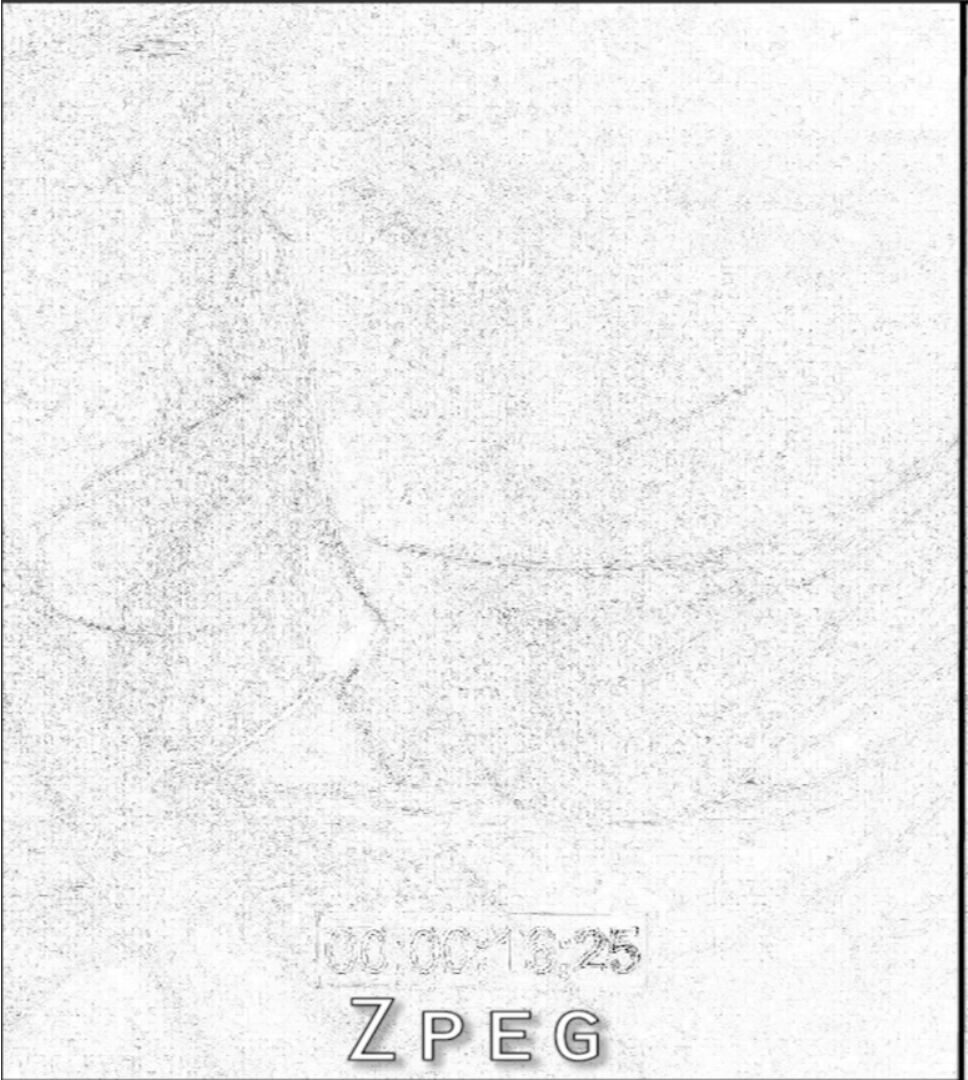
Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.

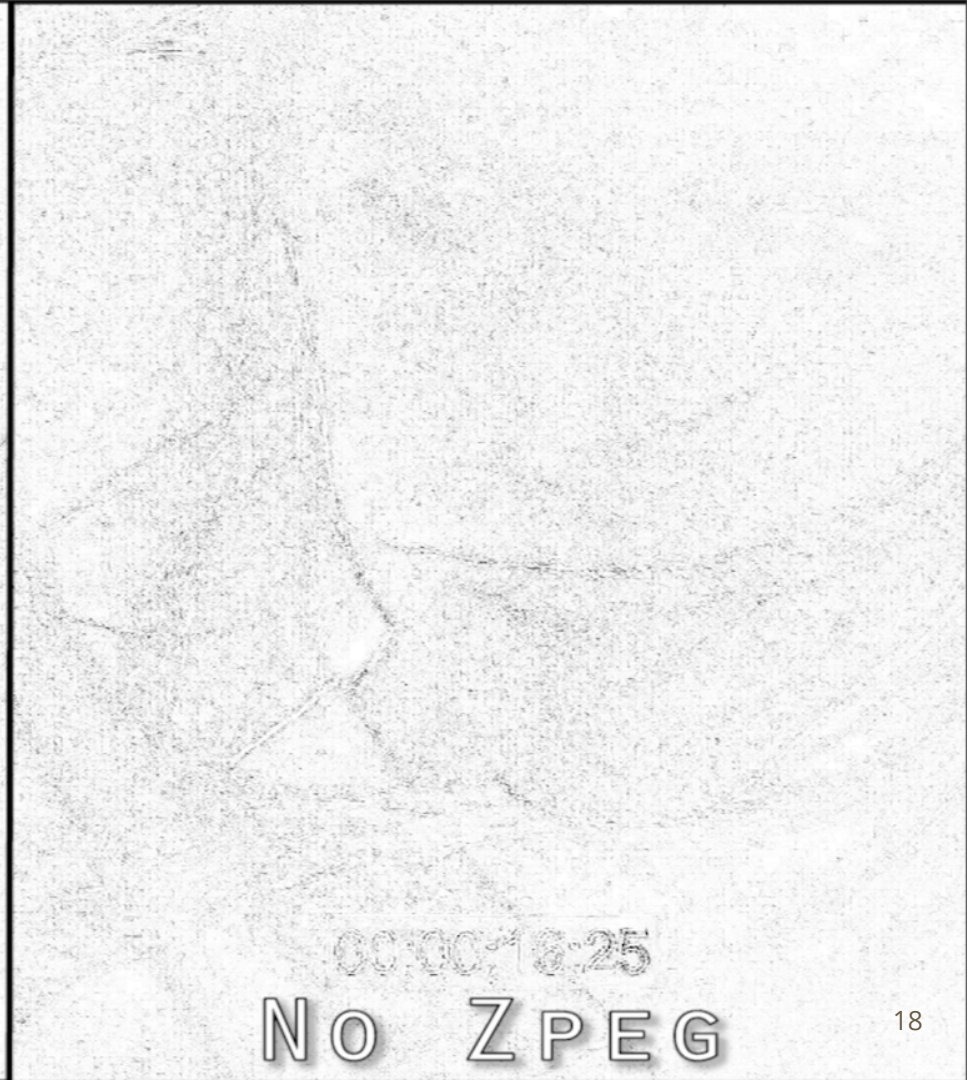

00;00;16;25
ZPEG
00;00;16;25
NO ZPEG

Quality Map View in SQM

This is the quality map view for the Elektra video. As you can see on the right (and in full screen on the following page), the difference maps are virtually identical, confirming the absence of artifacts that would be noticed by real time viewers.




00:00:18:25
ZPEG


00:00:18:25
NO ZPEG

Elektra

PSNR scan: no major qualitative differences;

Top/Bottom analysis: See video. Showed a reduction in film grain which may not be perceived as beneficial by creative types. Would not be noticed by normal viewers during real-time playback.

Note drop off at end: On several test clips, we noticed a quality drop at the very end resulting from a re-ordering of the last few frames. This is the sharp downward stroke of the red trace at the very right on top (clearer on the following page). We reported this in the first draft of this report. ZPEG sent a fix that we evaluate on the following page.

Bug Fix

Evaluation: We tested the fix on two clips, and found that it resolved the problem with no other observable issues. In the figure, the red trace is the original encode, the green the fix.

Otherwise, both files were almost exactly the same size as the previous file, though the PSNR score was slightly higher (under 1.5%) higher, indicating slightly higher quality.

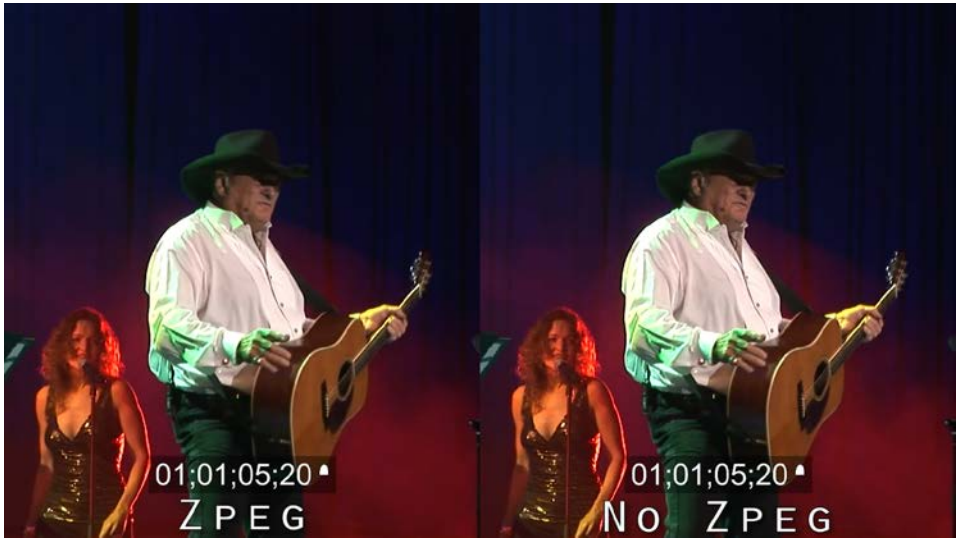
Since the report was otherwise almost complete, SLC decided to go ahead with the original tests and this explanation.

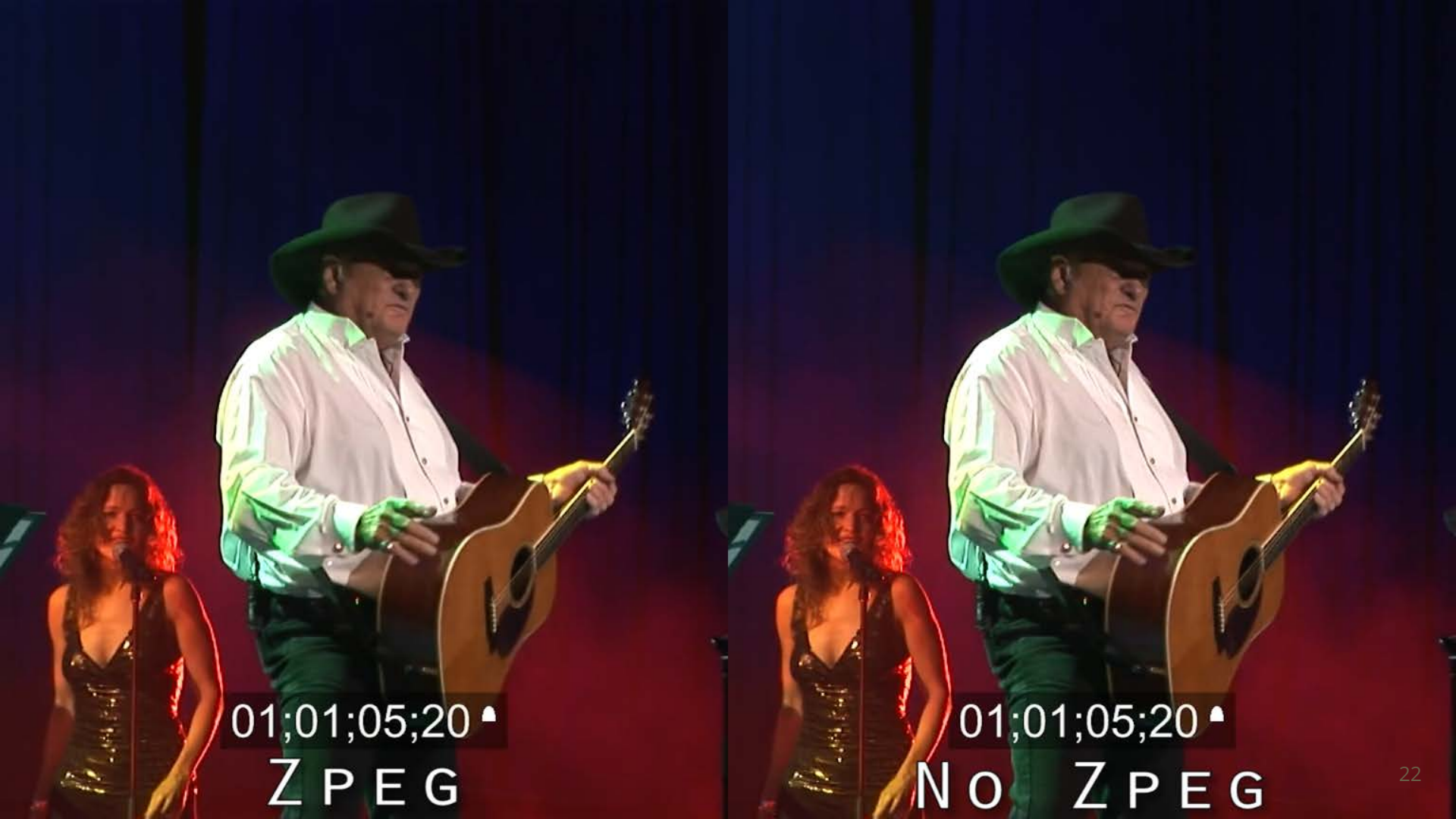
Freedom

Description: Freedom is a concert video of Josiah Weaver shot with a mix of HDV and AVCHD cameras.

Data rate reduction: -18.87%
VMAF differential: -3.72
PSNR differential: -1.47

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.

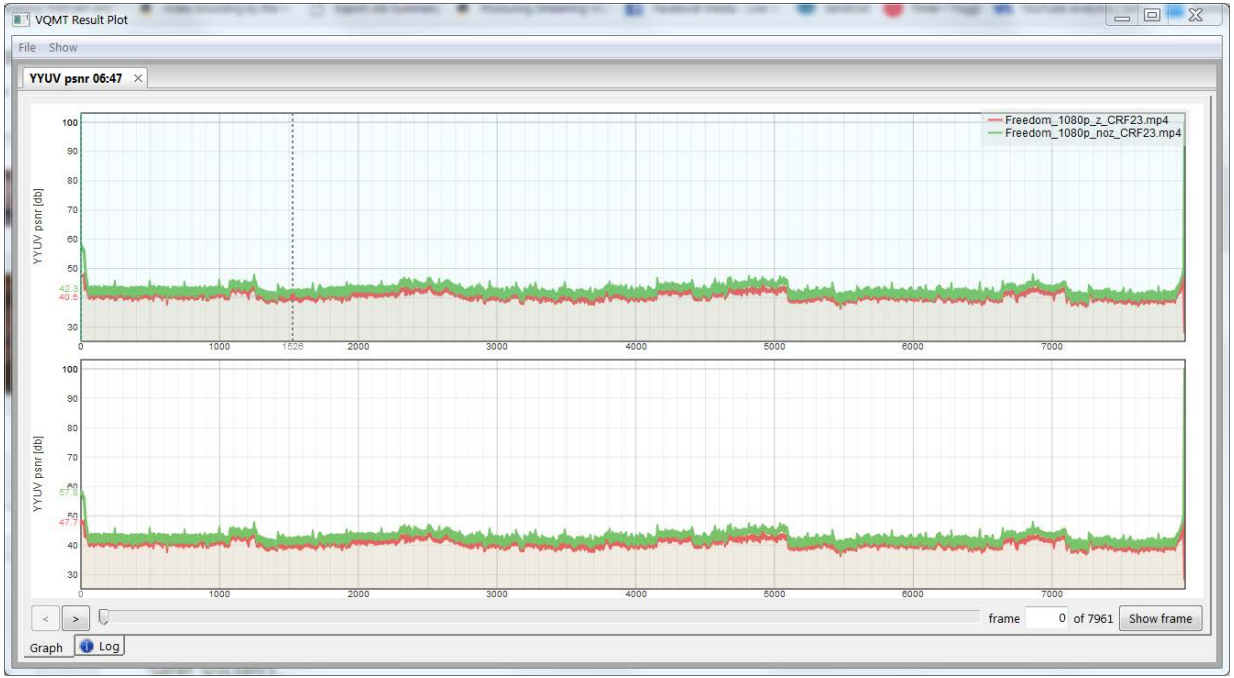



01;01;05;20
ZPEG
01;01;05;20
NO ZPEG

Freedom

PSNR scan: no major qualitative differences; On this demanding clip, the PSNR differential was minimal, which was impressive.

Top/Bottom analysis: See video. Some loss of detail (to be expected, and not noticeable by real time viewers).

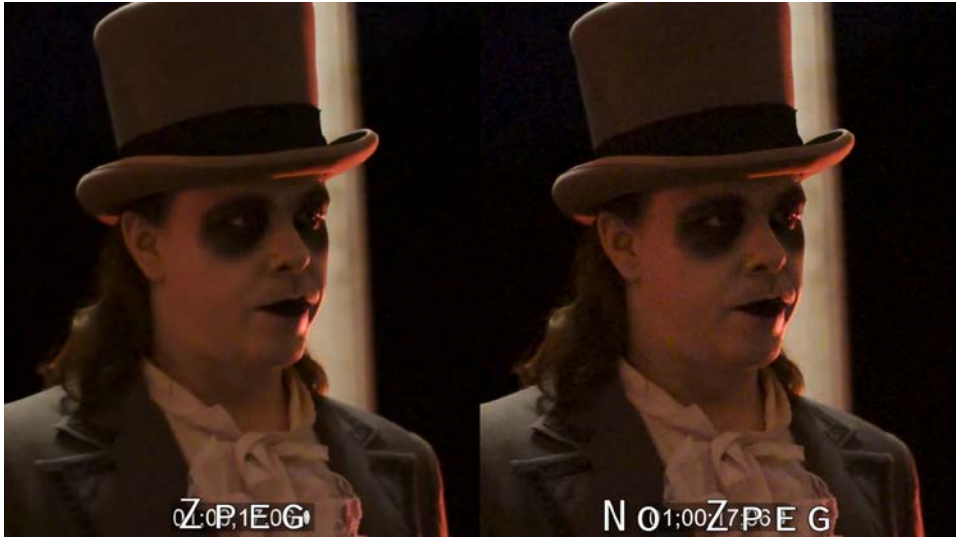


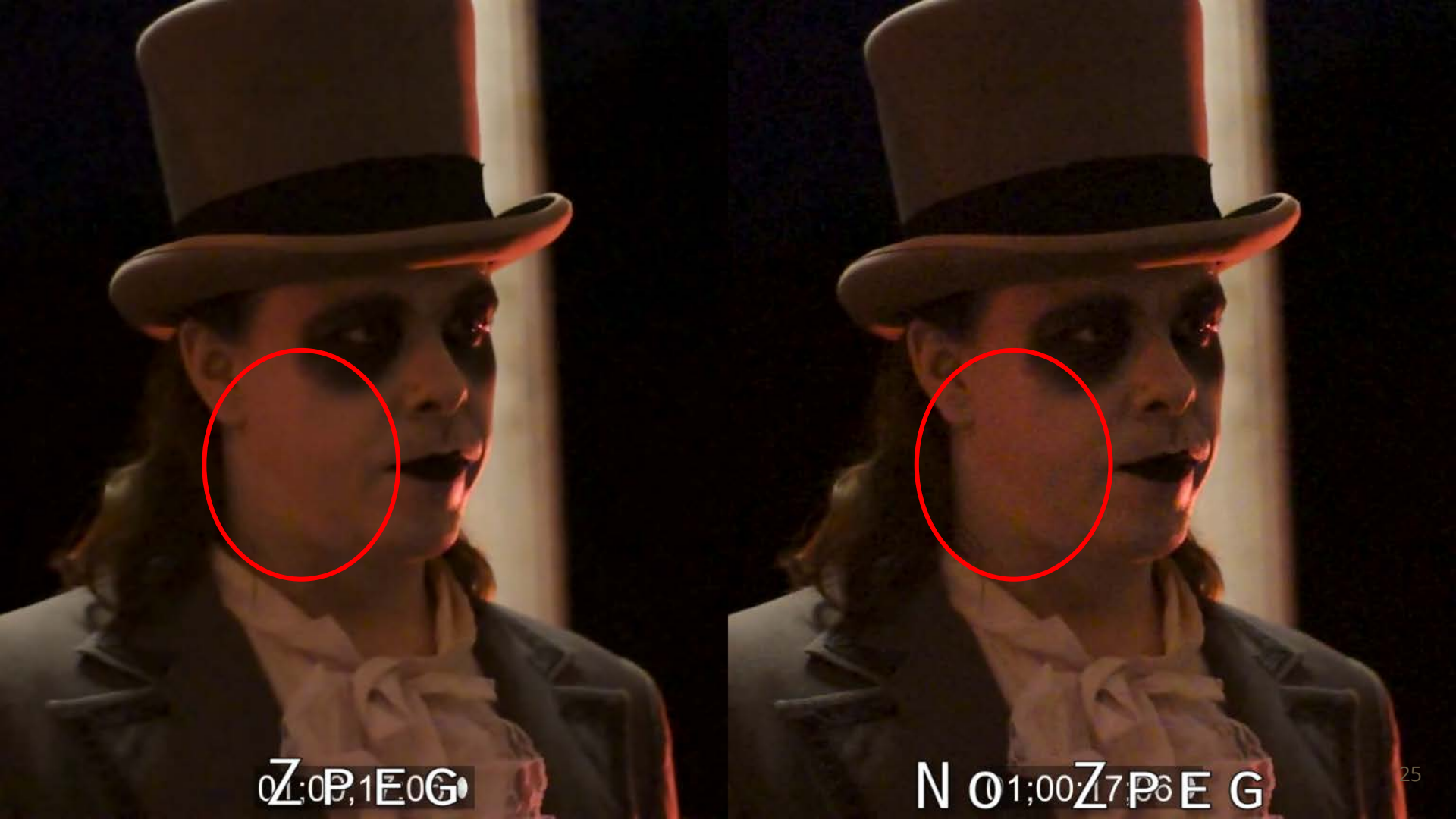
Haunted

Description: Haunted is a faux-movie-like production advertising a haunted house. Shot with a DSL in awful conditions which created lots of noise in the video.

Data rate reduction: -33.02%
VMAF differential: -5.93
PSNR differential: -1.45

Subjective quality analysis: No visual difference in real-time. Slightly less artifacts in frame by frame; not noticeable by real-time viewer.

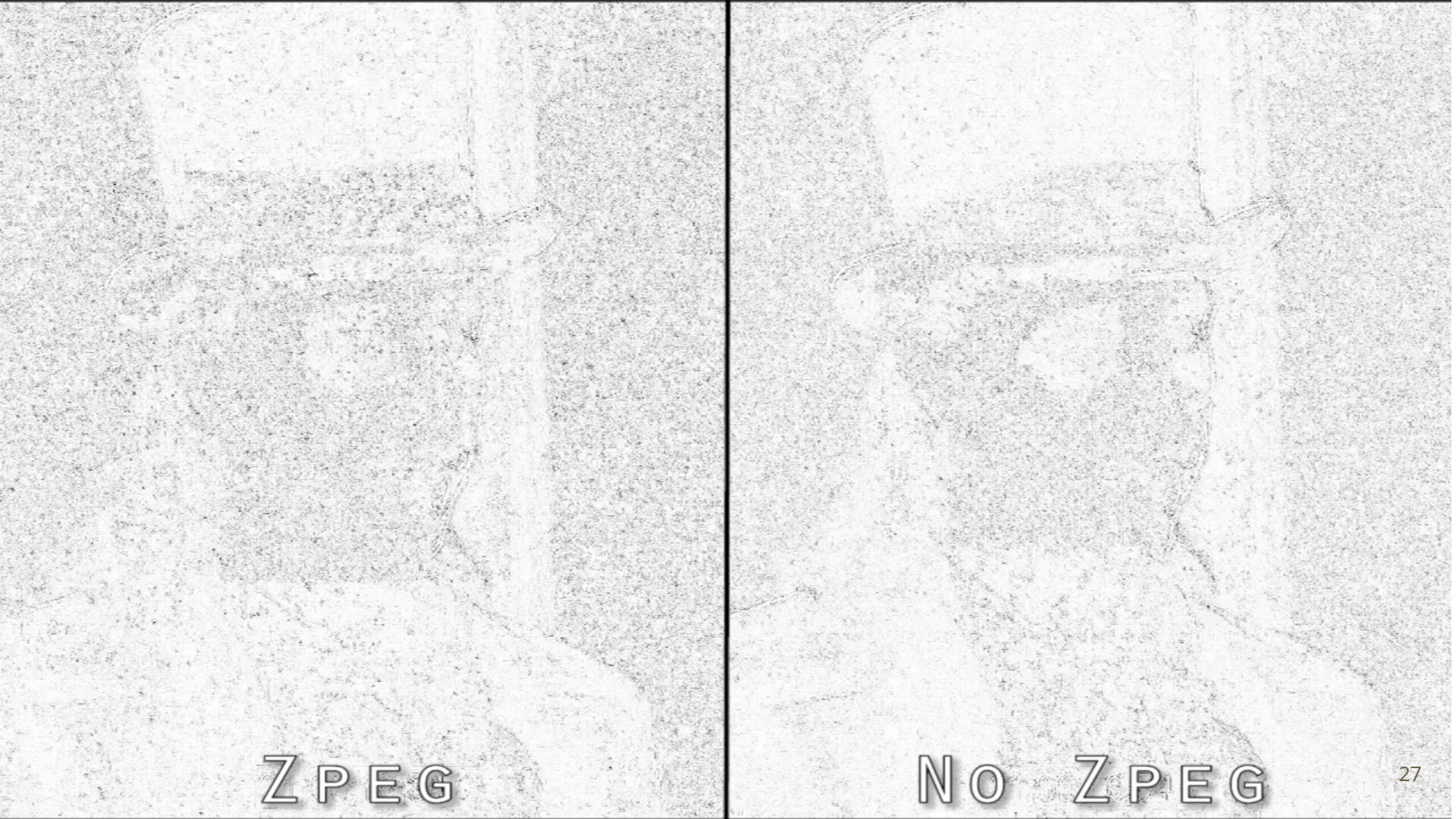



ZIP JPEG
NO ZIP JPEG

Quality Map View in SQM

This is the quality map view for the Haunted video. As you can see on the right (and in full screen on the following page), the difference maps are similar, except that it appears as if ZPEG is removing slightly more noise from the face and background.

In this case, the noise reduction quality actually improved the overall frame by removing some minor artifacts, though this wouldn't be noticeable by a viewer without side-by-side displays or top/bottom analysis.

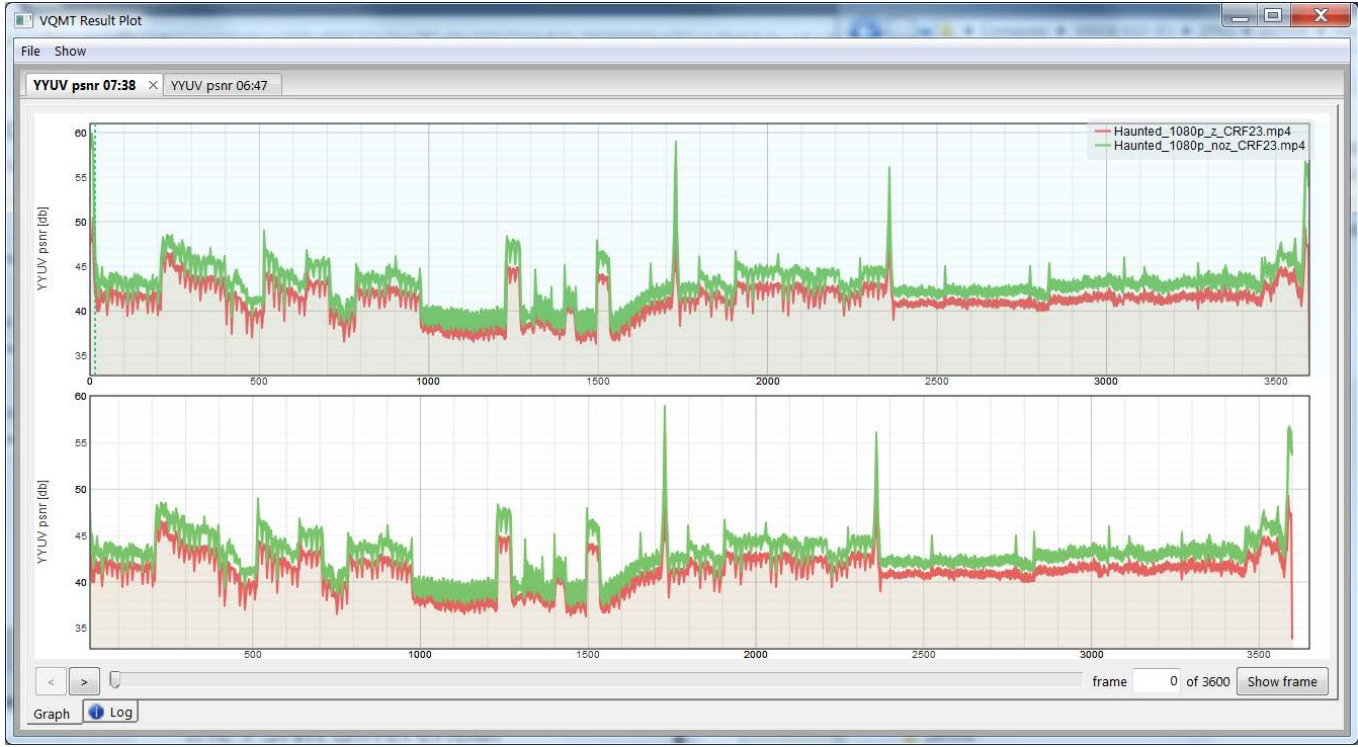

ZPEG
NO ZPEG

Haunted

PSNR scan: no major qualitative differences; As with the Freedom clip, the PSNR differential was minimal, which was impressive on this noisy clip.

Top/Bottom analysis: See video. Showed lots of noise removal plus a reduction in vibrance/brightness in one scene. Would not be noticed by normal viewers during real-time playback.

You can see the downward spike on the extreme right of the bottom window. This is the previously reported bug.



Bug Fix

Results: This was the other clip we tested with the bug fix (in green) against the original encode (in red), which showed the drop in quality due to frame reordering at the end.

As you can see on the right, the fix did resolve the problem, and in this case produced a PSNR score of about 1% higher than the original with no other observable issues.

Zoolander

Description: Zoolander is the first five minutes from the first Zoolander movie. Both VMAF and PSNR differentials are impressive given the significant reduction in data rate.

Data rate reduction: -27.96%
VMAF differential: -2.81
PSNR differential: -1.81

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.


00:03:59:10
ZPEG
00:03:59:10
NO ZPEG

Zoolander

PSNR scan: no major qualitative differences; As with the Freedom clip, the PSNR differential was minimal, which was impressive on this noisy clip.

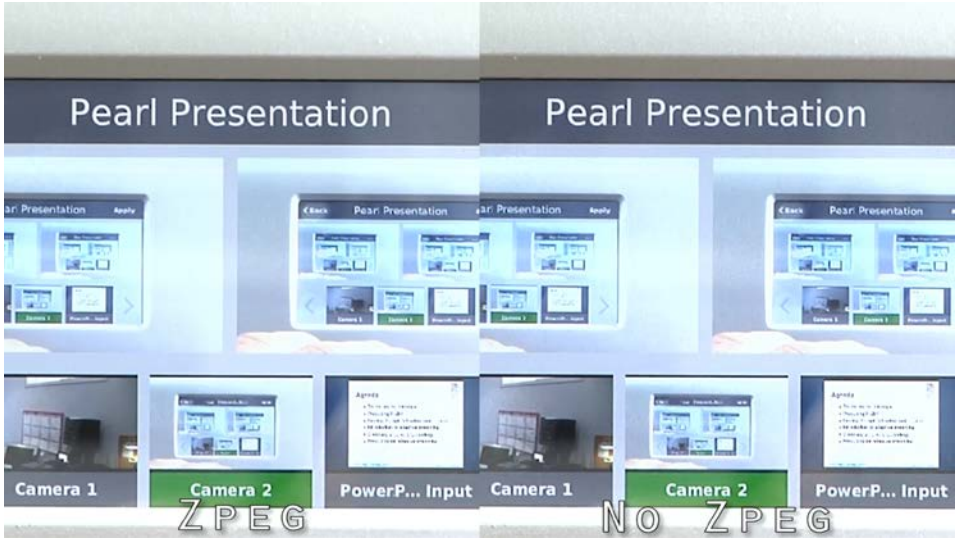
Top/Bottom analysis: See video. Showed lots of noise reduction plus a slight reduction in vibrance/ brightness in one scene. Would not be noticed by normal viewers during real-time playback.

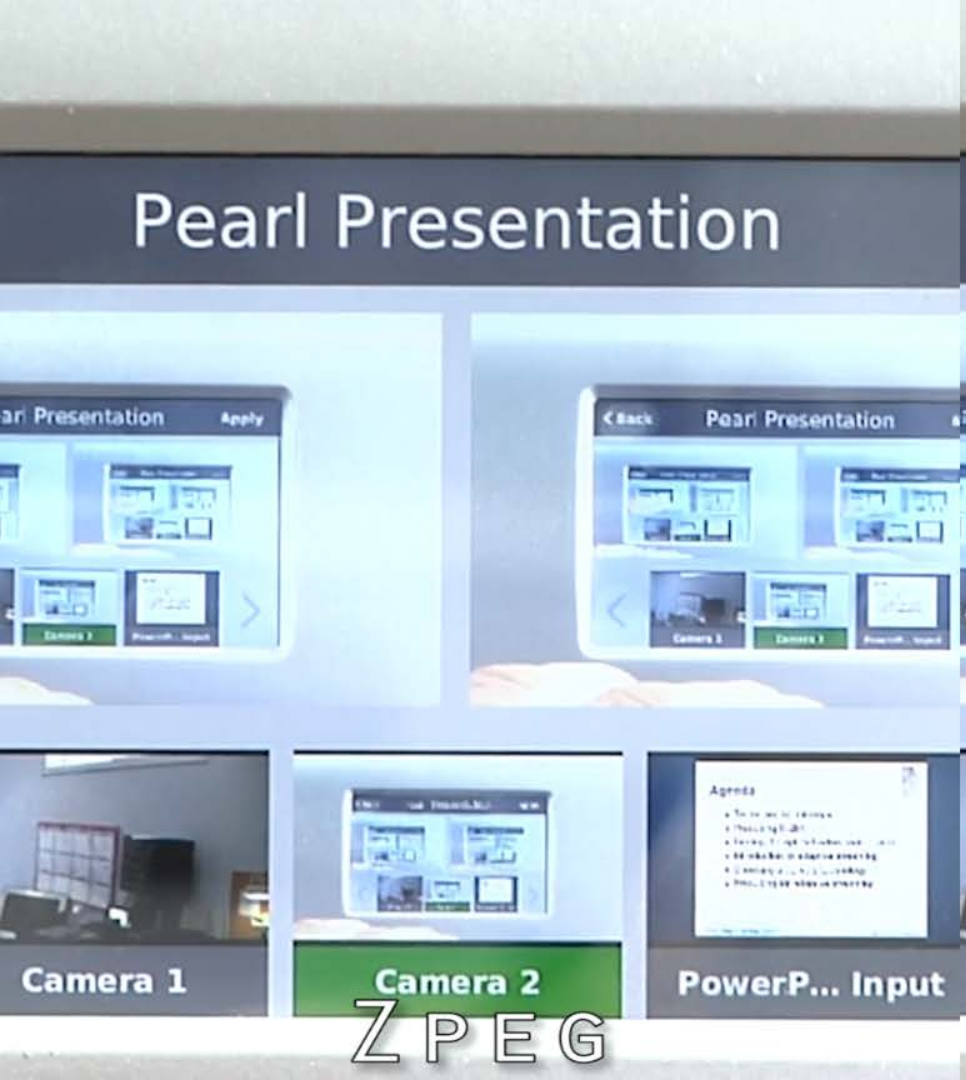
Epiphan

Description: Epiphan is a video tutorial consisting of multiple real world shots (no screencams). It's a mix of instructional (on right) and talking head. It was shot with an AVCHD camcorder under good conditions.

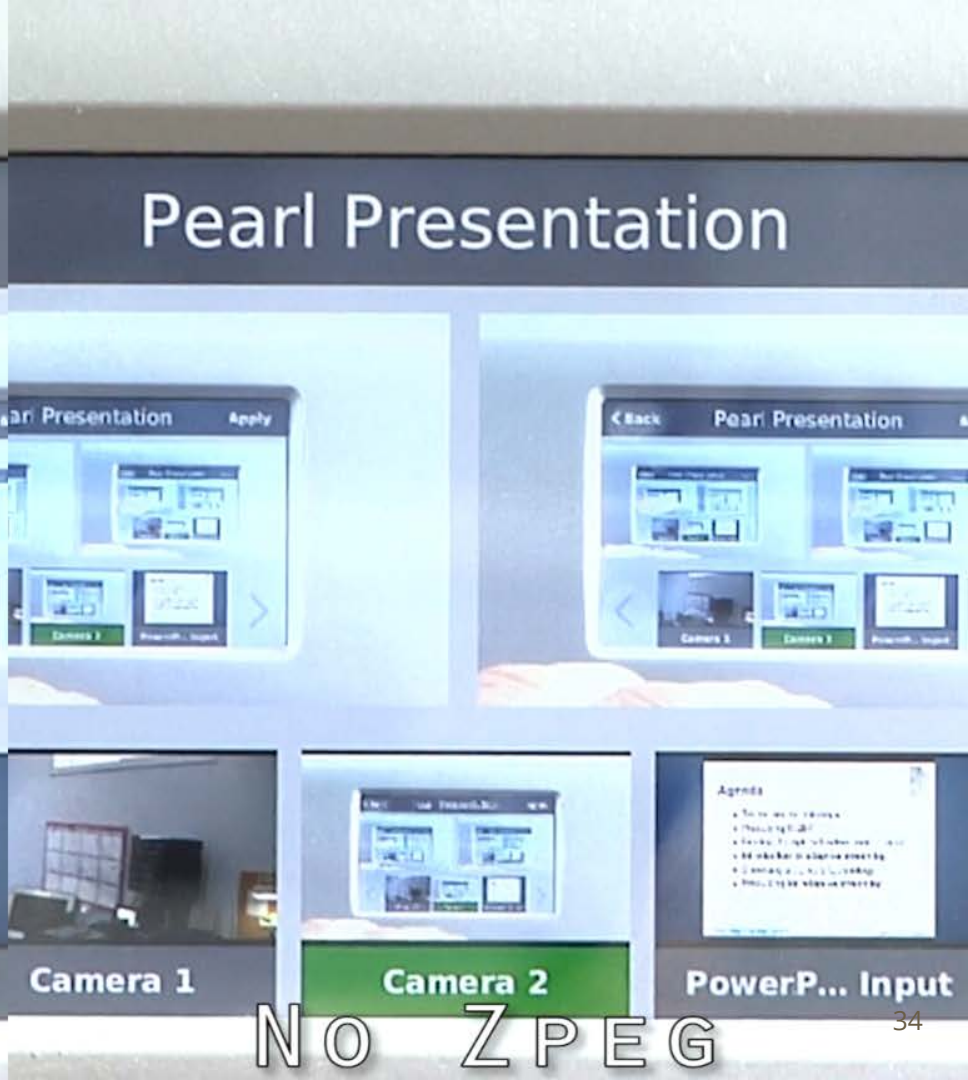
Data rate reduction: -38.11%
VMAF differential: -2.76
PSNR differential: -2.53

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.




Pearl Presentation
Camera 1
Camera 2
PowerP... Input

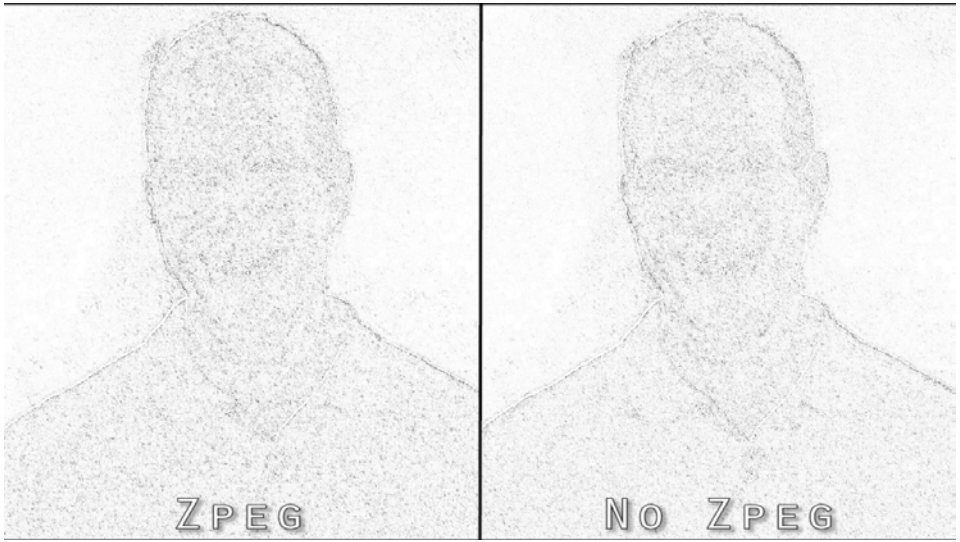
ZPEG


Pearl Presentation
Camera 1
Camera 2
PowerP... Input

NO ZPEG

Quality Map View in SQM

This is the quality map view for the Epiphan video. As you can see on the right (and in full screen on the following page), the difference maps are similar, except that it appears as if ZPEG is removing slightly more noise from the face and shirt, not surprising given the impressive 38% reduction in file size.




ZPEG
NO ZPEG

Epiphan

PSNR scan: no major qualitative differences; Where the NoZ clip soars to 100 on the left, it's a black frame and there's no visible difference between the Z and NoZ frames.

Top/Bottom analysis: See video. Good noise reduction, with a slight dulling of skin tone. Would not be noticed by normal viewers during real-time playback.

New

Description: An amalgam of nine test videos shot by a Red camera. All very good quality and reasonably well lit.

Data rate reduction: -17.86%
VMAF differential: -3.90
PSNR differential: -1.72

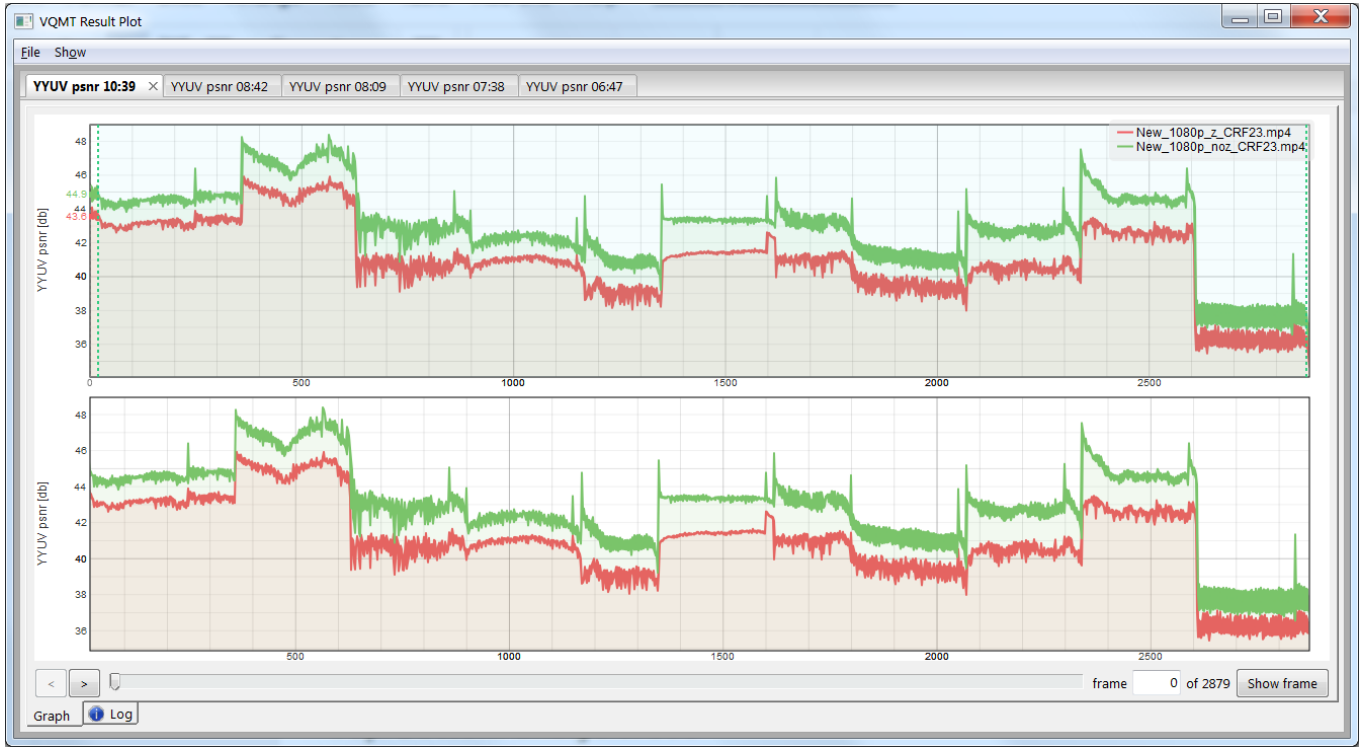
Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.


FRANZIA
MIST
Double Gold Medal

New

PSNR scan: no major qualitative differences; overall differential of 1.72 is even across the entire clip. Even when low on the extreme right, no differences visible in real time.

Top/Bottom analysis: See video. Slight color adjustments; whites are very accurate, some colors not so much. Nothing anyone would notice outside of a T/B view.

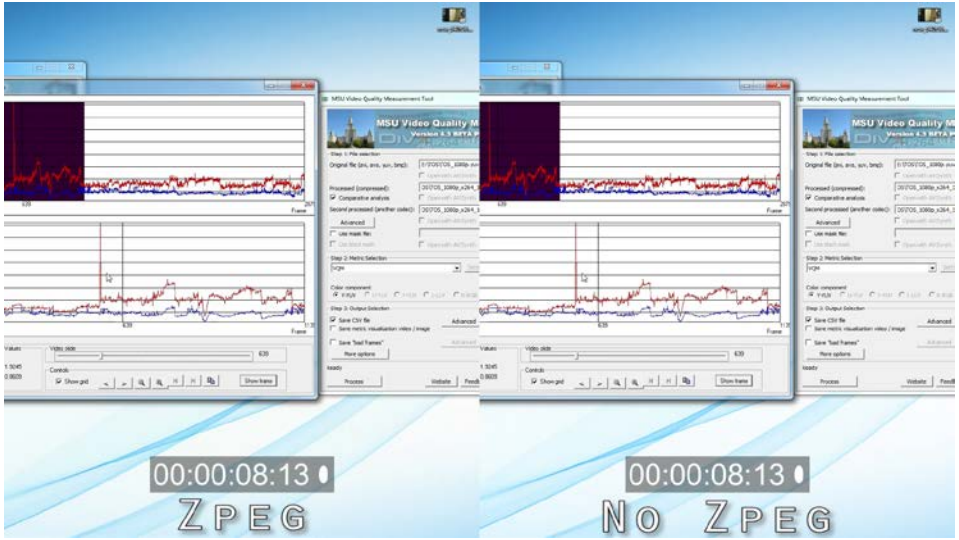


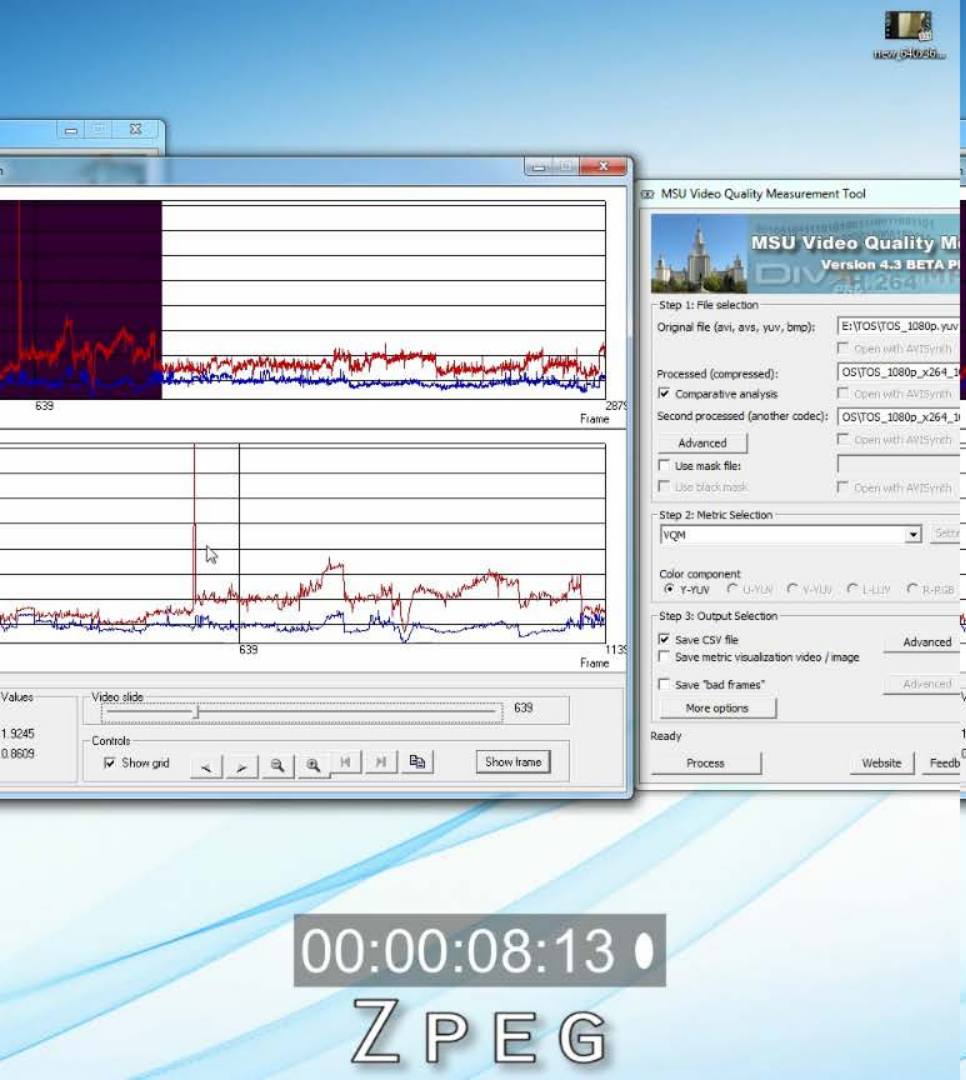
Screencam

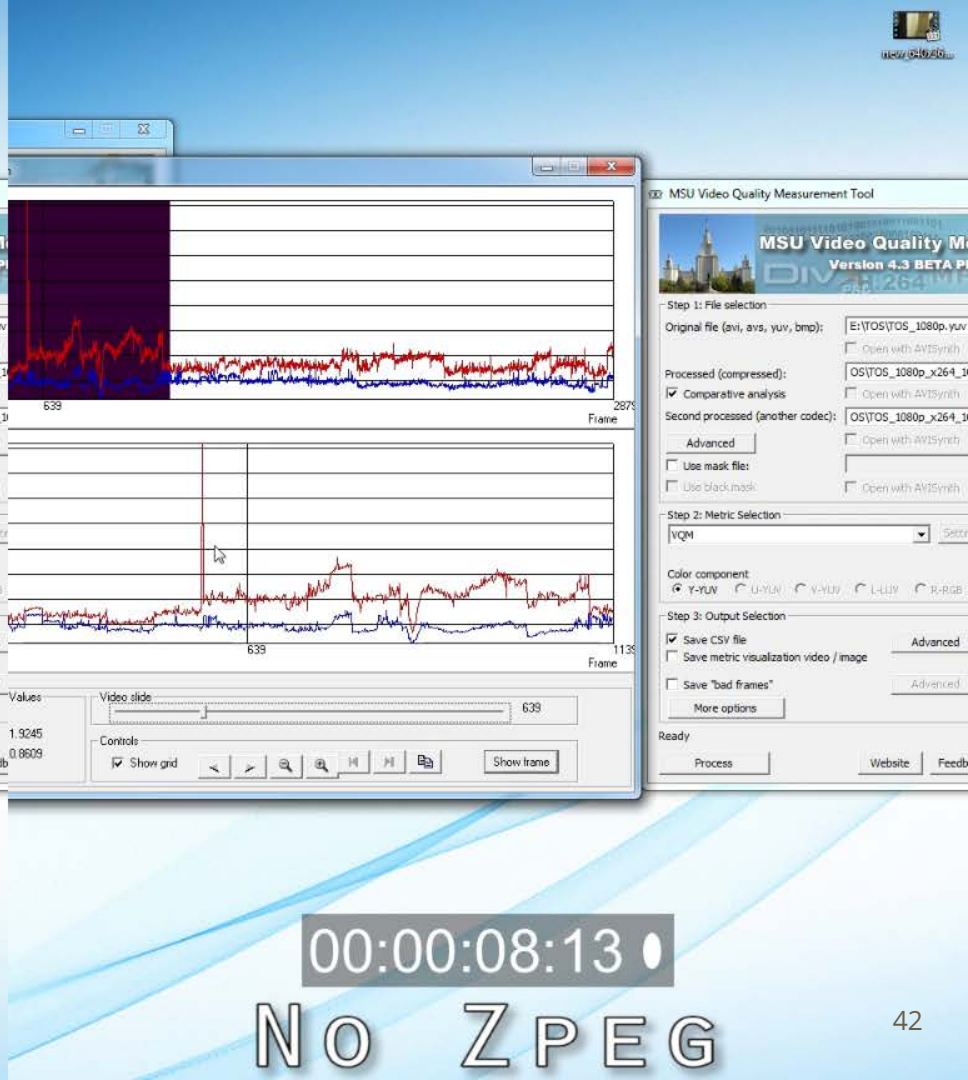
Description: A screencam review of the Moscow State University Video Quality Metric showing both screens (on right) and some video frames (not shown). Minimal data rate reduction, almost no differential in any views. Would not use ZPEG for screencams.

Data rate reduction: -5.97%
VMAF differential: -2.10
PSNR differential: -3.04

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.




MSU Video Quality Measurement Tool
MSU Video Quality M
Version 4.3 BETA P
Step 1: File selection
Original file (avi, avs, yuv, bmp):
E:\TOS\TOS_1080p.yuv
Open with AVISynth
Processed (compressed):
OS\TOS_1080p_x264_1
Comparative analysis
Second processed (another codec):
Advanced
Use mask file:
Use black mask
Step 2: Metric Selection
VQM
Color component
Y-YUV
Step 3: Output Selection
Save CSV file
Save metric visualization video / image
Save "bad frames"
More options
Ready
Process
Website
Feedb
639
Frame
Video slide
Controls
Show grid
Show frame
00:00:08:13
ZPEG


MSU Video Quality Measurement Tool
MSU Video Quality M
Version 4.3 BETA P
Step 1: File selection
Original file (avi, avs, yuv, bmp):
E:\TOS\TOS_1080p.yuv
Open with AVISynth
Processed (compressed):
OS\TOS_1080p_x264_1
Comparative analysis
Second processed (another codec):
Advanced
Use mask file:
Use black mask
Step 2: Metric Selection
VQM
Color component
Y-YUV
Step 3: Output Selection
Save CSV file
Save metric visualization video / image
Save "bad frames"
More options
Ready
Process
Website
Feedb
639
Frame
Video slide
Controls
Show grid
Show frame
00:00:08:13
NO ZPEG

Screencam

PSNR scan: Extreme differences relate to dropped frames in both CRF clips. They recover after a single frame; issue would not be visible.

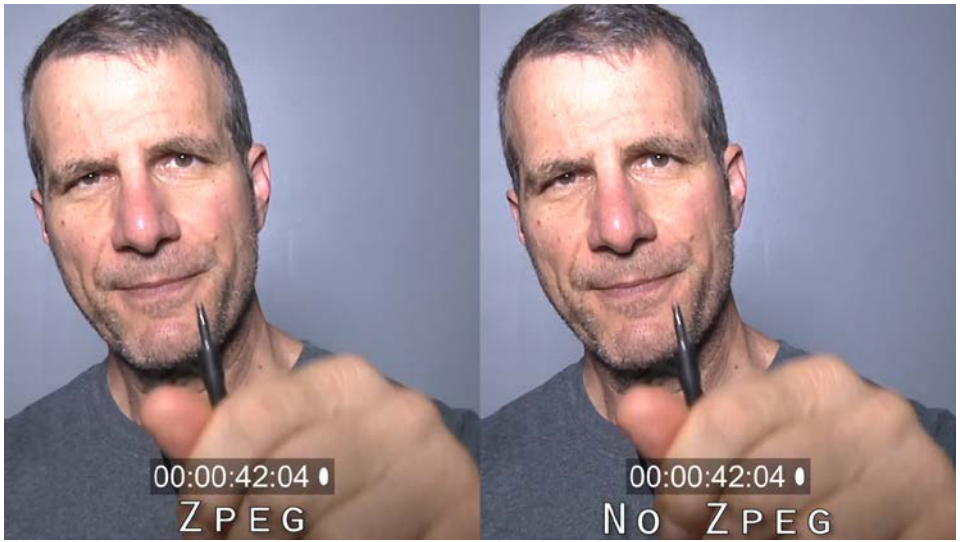
Top/Bottom analysis: No differences seen.

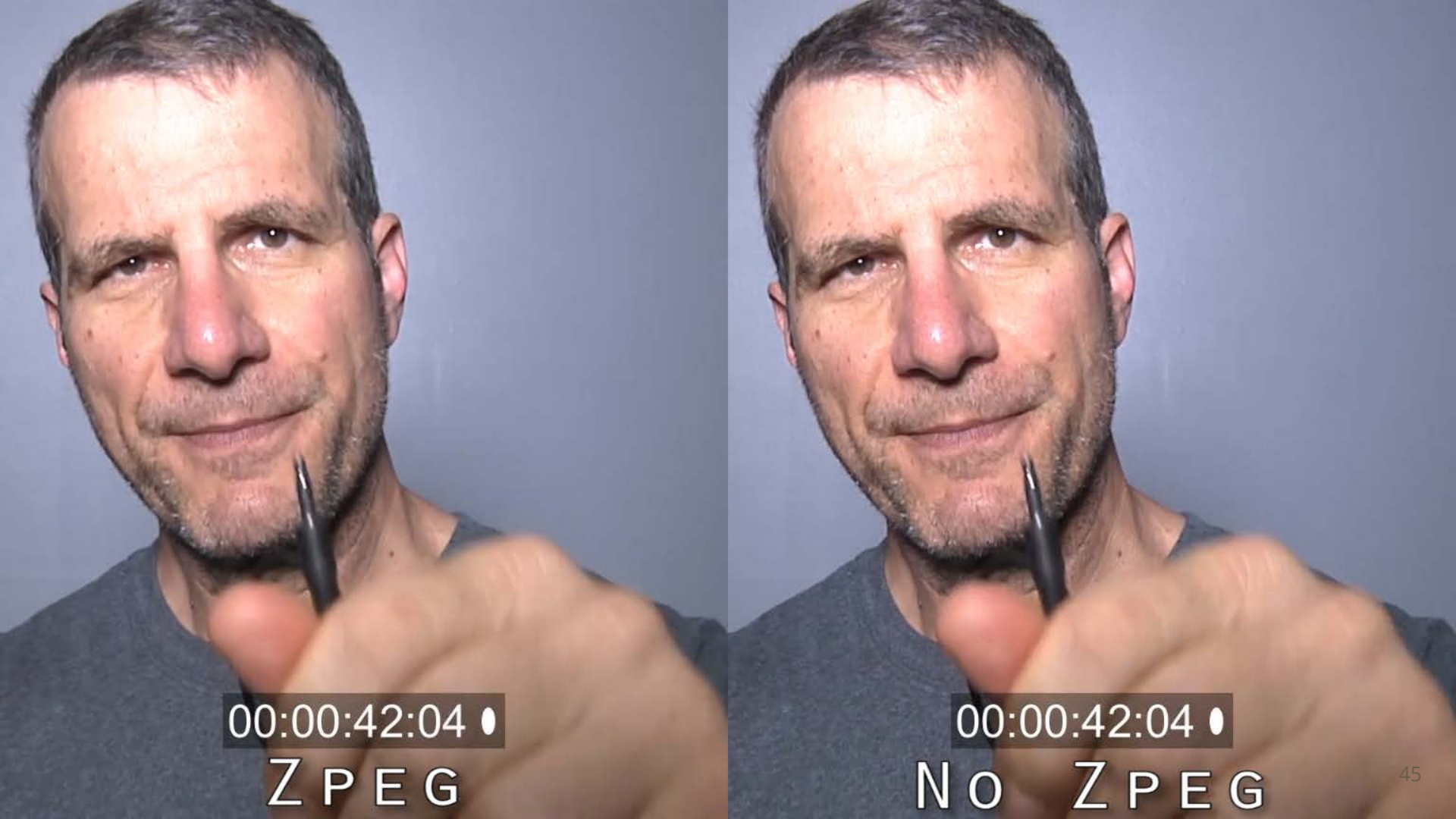
Talking Head

Description: This is a simple talking head video from a video review of a Sennheiser wireless microphone system, shot with AVCHD in good lighting.

Data rate reduction: -38.36%
VMAF differential: -2.68
PSNR differential: -1.58

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.




00:00:42:04
ZPEG
00:00:42:04
NO ZPEG

Quality Map View in SQM

This is the quality map view for the talkinghead video. As you can see on the right (and in full screen on the following page), the difference maps are similar, except that it appears as if ZPEG is removing slightly more noise from the face and around the head, which is appropriate given the impressive 38% reduction in file size.

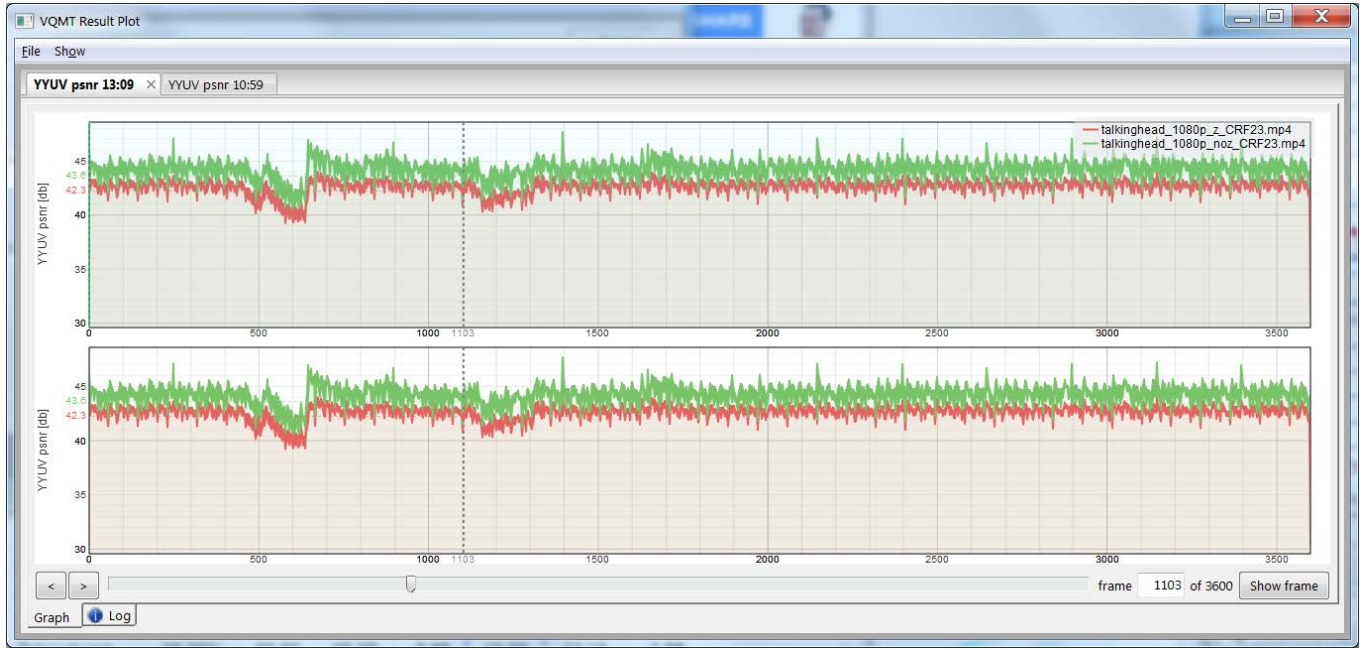
No blocks, jaggies, or other artifacts shown, making it unlikely that a viewer watching in real time would notice the difference.


ZPEG
NO ZPEG

Talking Head

PSNR scan: No major differences in the scan.

Top/Bottom analysis: ZPEG retained very good color fidelity here, with some very minor loss of detail as you would expect with a data rate reduction of 38%. Performance here was a real surprise for such a simple clip. Check the video for more details.

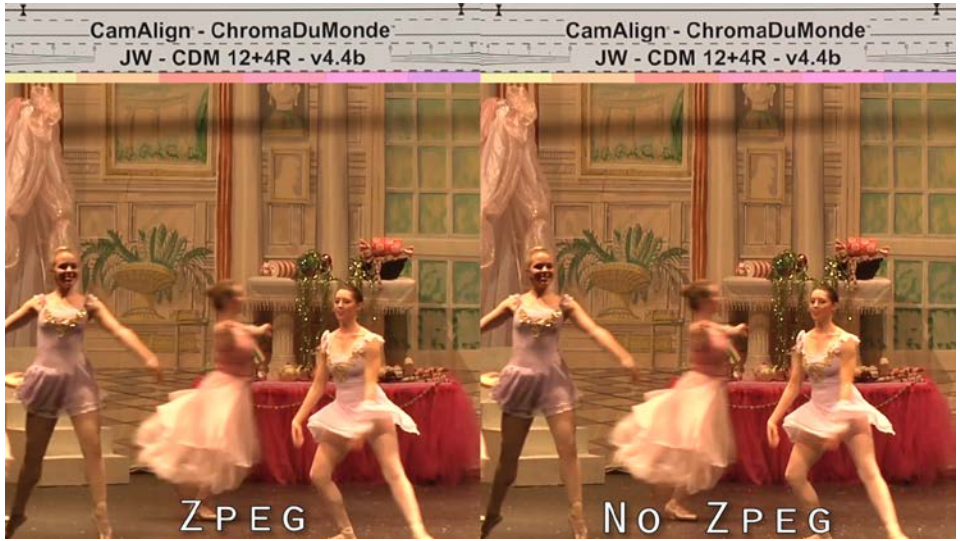


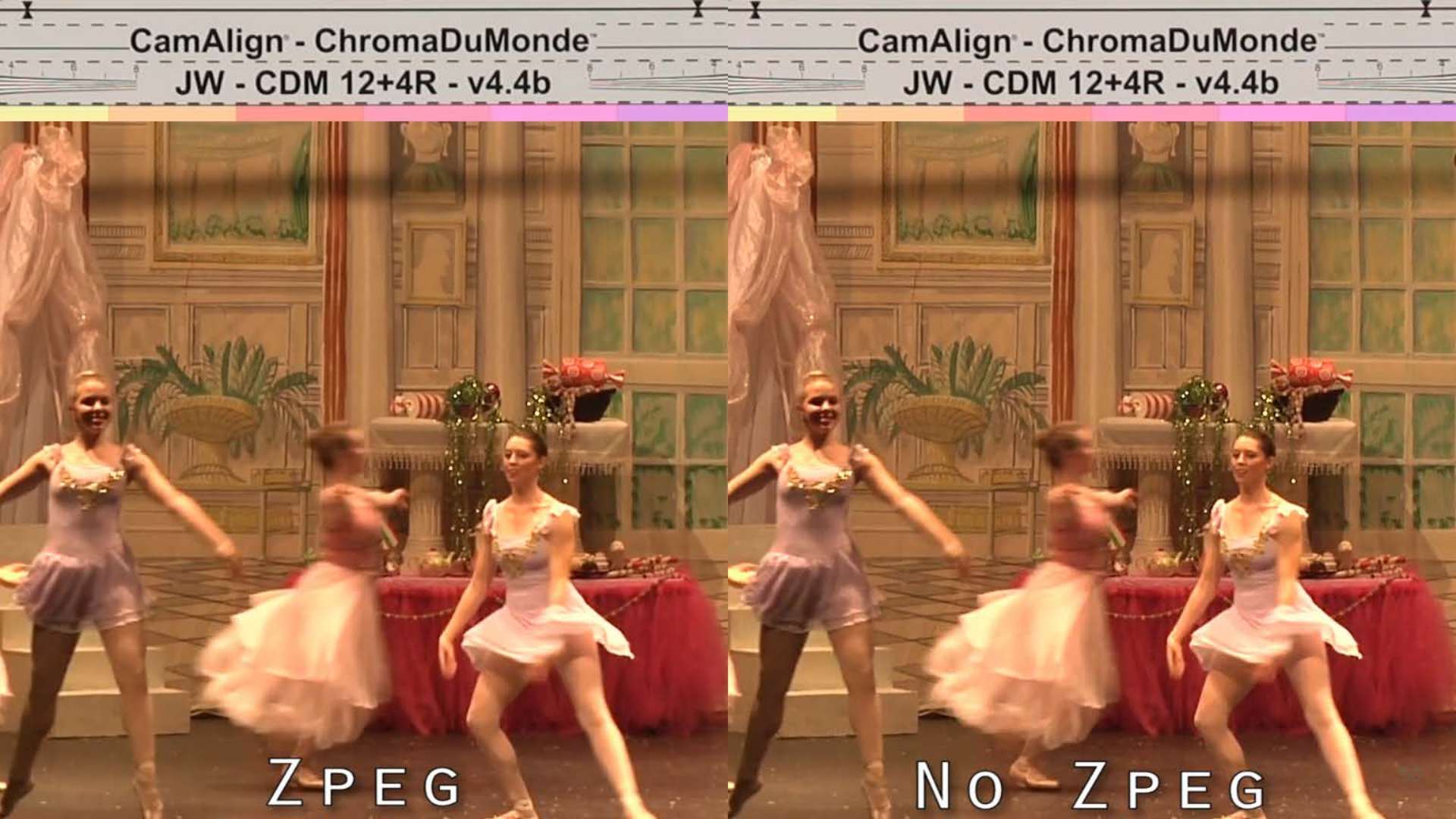
Test

Description: This is an 8-minute video comprised of 30 seconds of talking head from the previous clip and 30 seconds of ballet. Originally created to test bitrate control techniques like CBR, constrained VBR and capped CRF.

Data rate reduction: -30.33%
VMAF differential: -3.07
PSNR differential: -1.18

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.

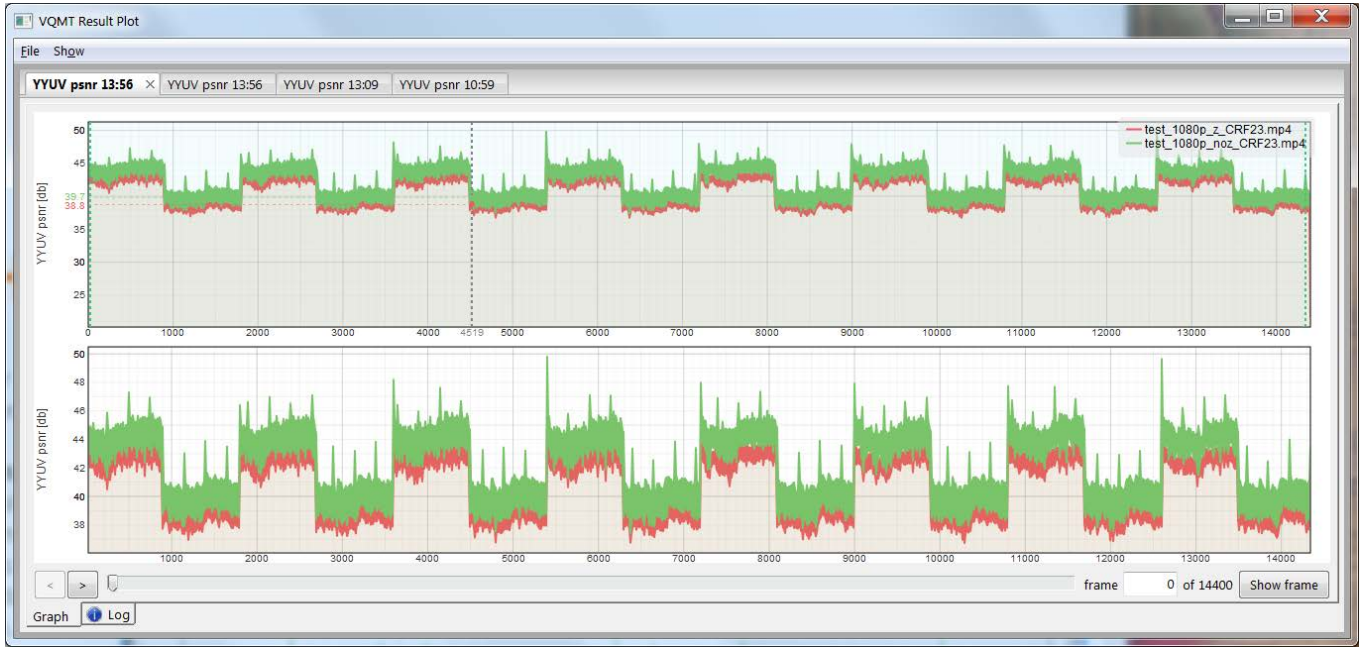



CamAlign - ChromaDuMonde
JW - CDM 12+4R - v4.4b
CamAlign - ChromaDuMonde
JW - CDM 12+4R - v4.4b
ZPEG
NO ZPEG

Talking Head

PSNR scan: No major differences in the scan.

Top/Bottom analysis: Nothing noted. Good color fidelity and detail throughout

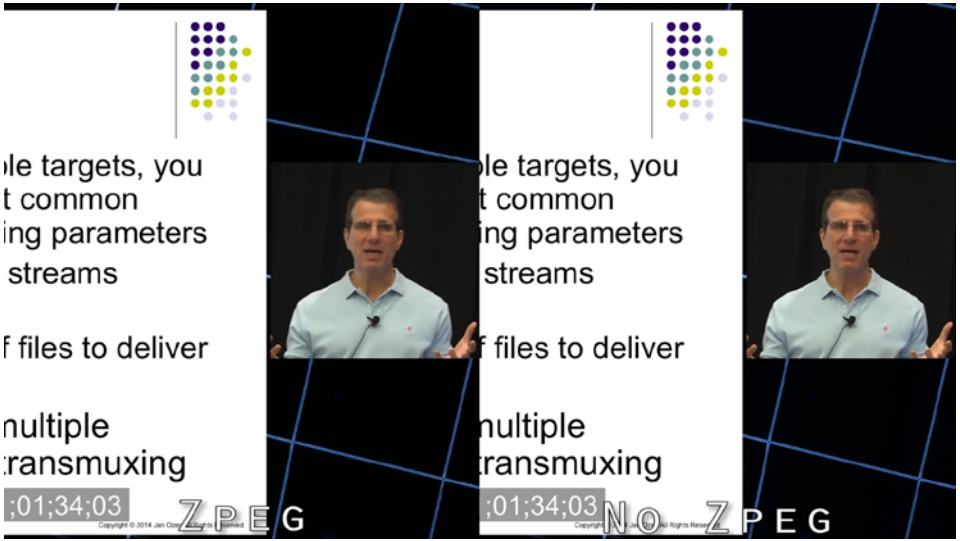


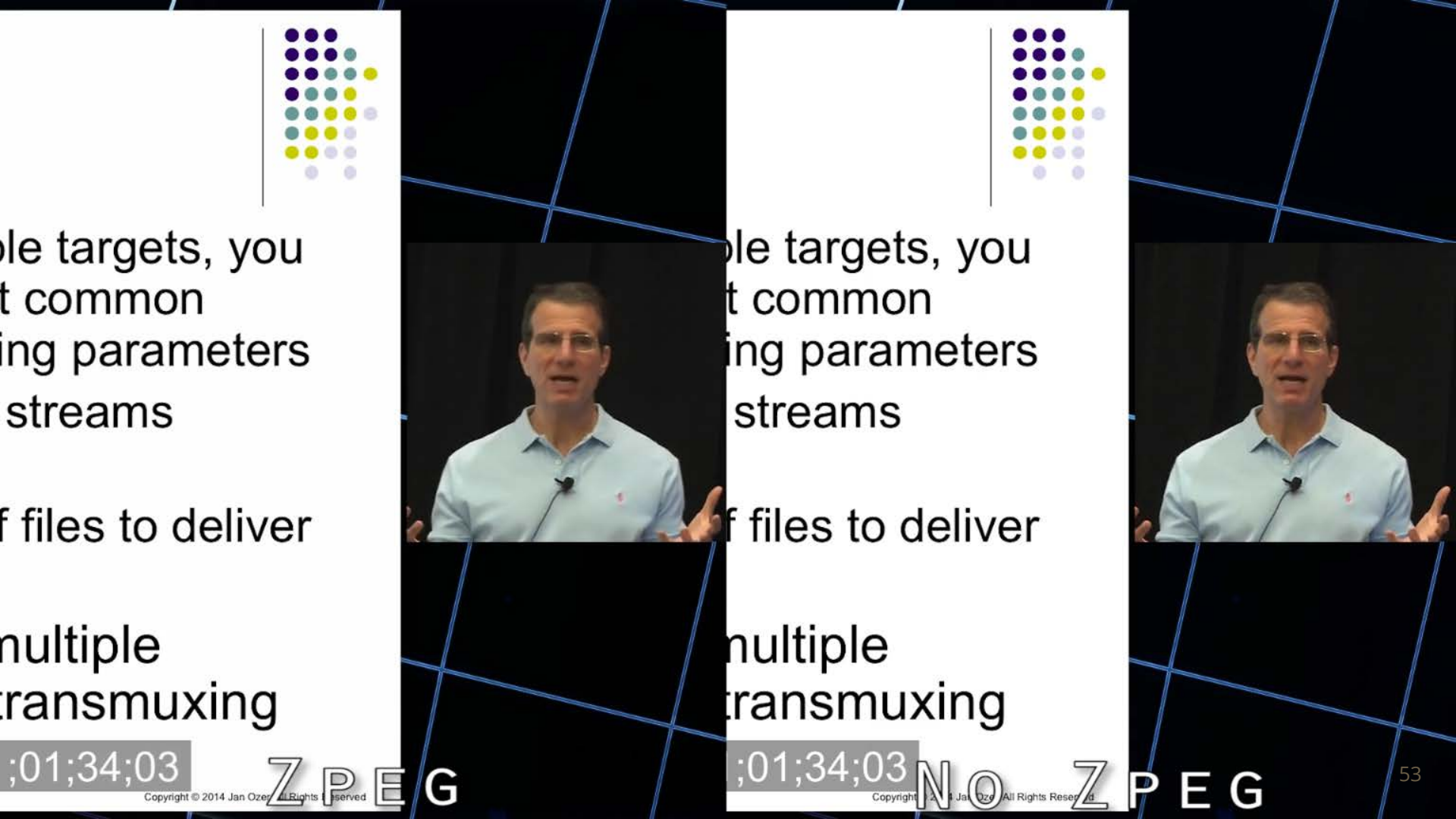
Tutorial

Description: This video is comprised of PowerPoint slides and a small talking head video on the right.

Data rate reduction: -17.39%
VMAF differential: -1.31
PSNR differential: -3.19

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.

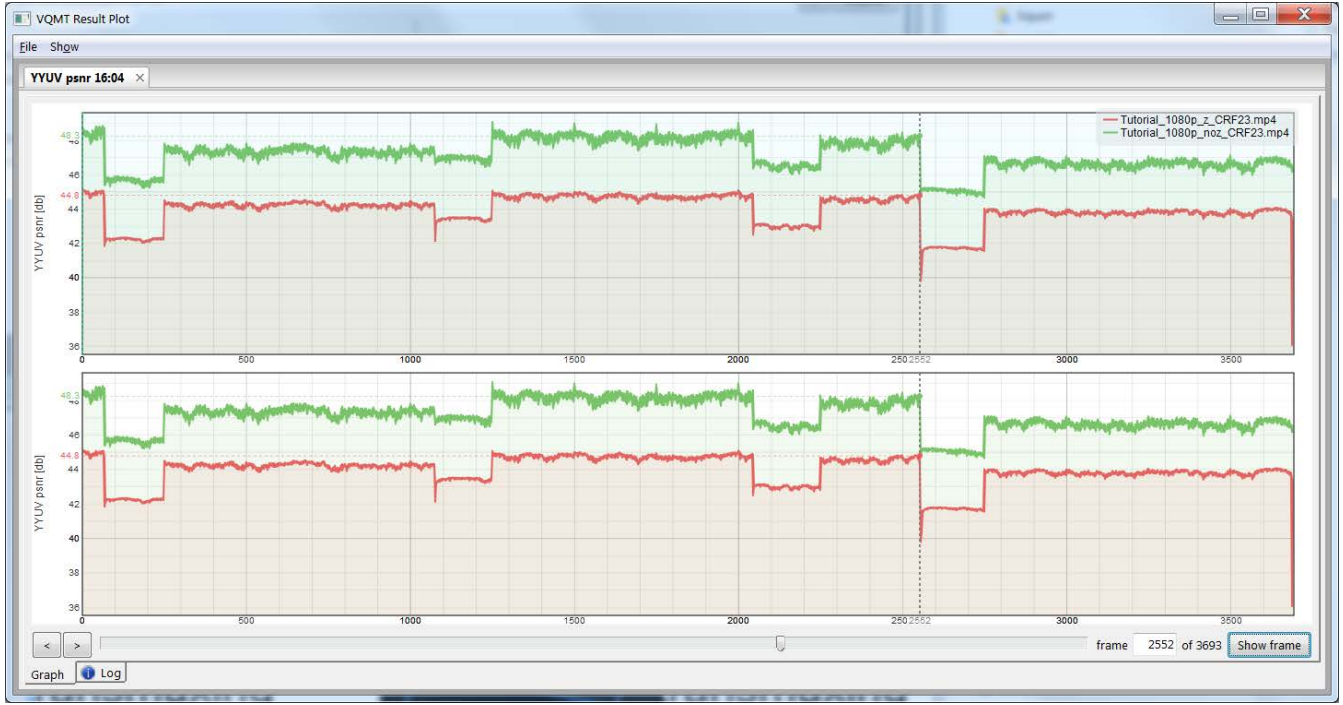



ble targets, you
t common
ing parameters
streams
f files to deliver
multiple
transmuxing
;01;34;03
Copyright © 2014 Jan Ozer All Rights Reserved
ZPEG
ble targets, you
t common
ing parameters
streams
f files to deliver
multiple
transmuxing
;01;34;03
Copyright © 2014 Jan Ozer All Rights Reserved
NO ZPEG

Talking Head

PSNR scan: No major differences in the scan.

Top/Bottom analysis: Nothing noted. Retained good color fidelity throughout.

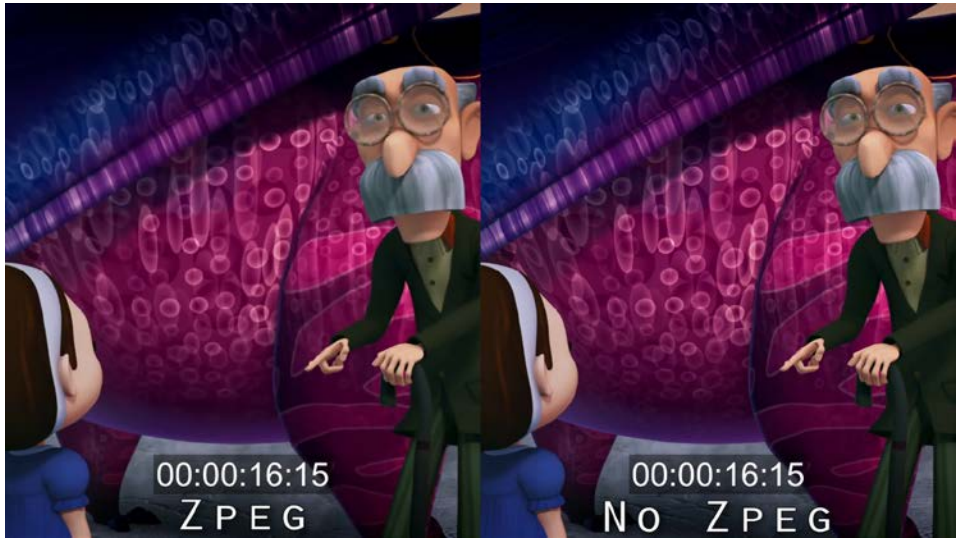


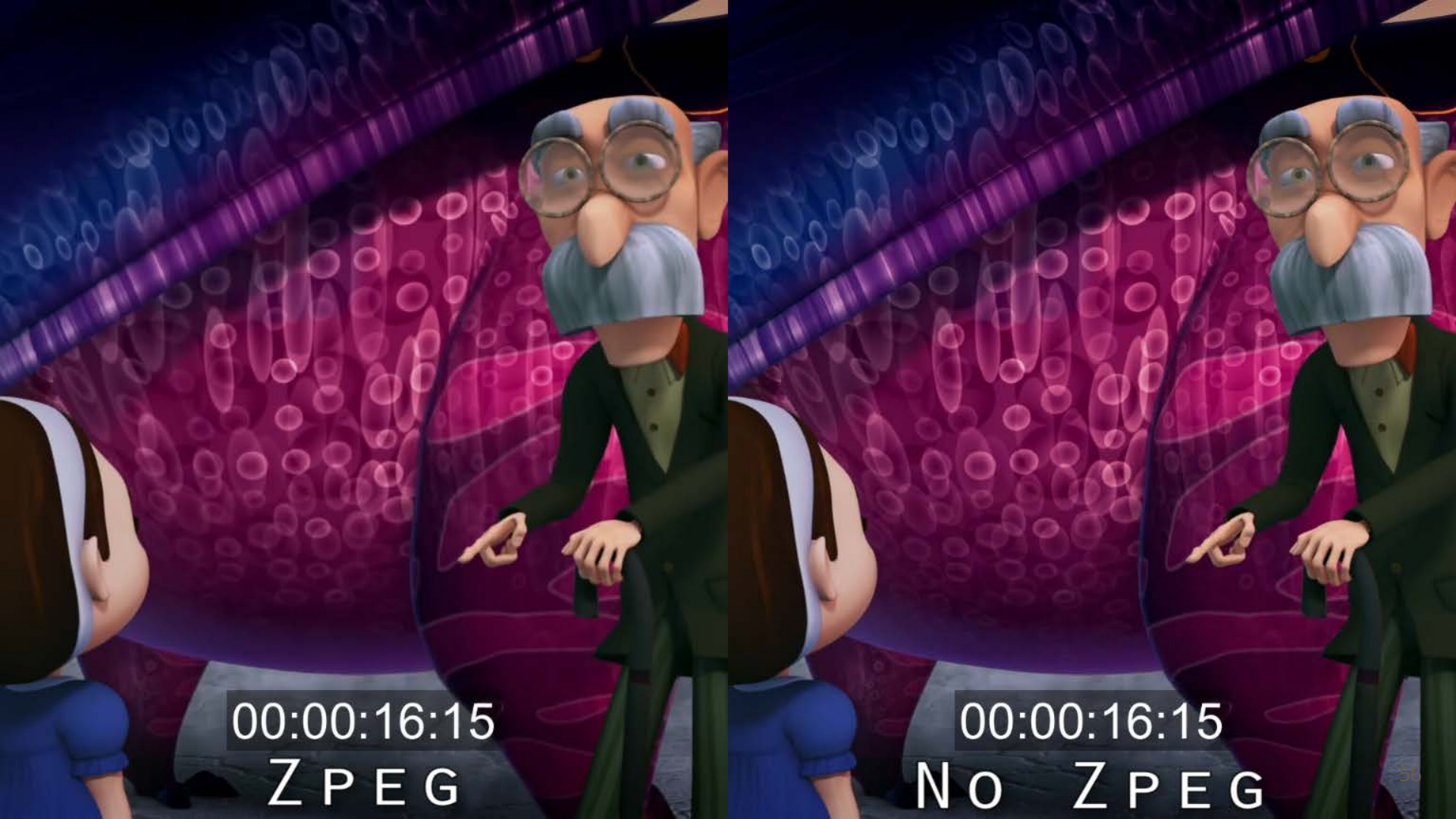
El-Ultimo

Description: El Ultimo is a simple 2D animated movie.

Data rate reduction: 9.83%
VMAF differential: -3.18
PSNR differential: -2.07

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.



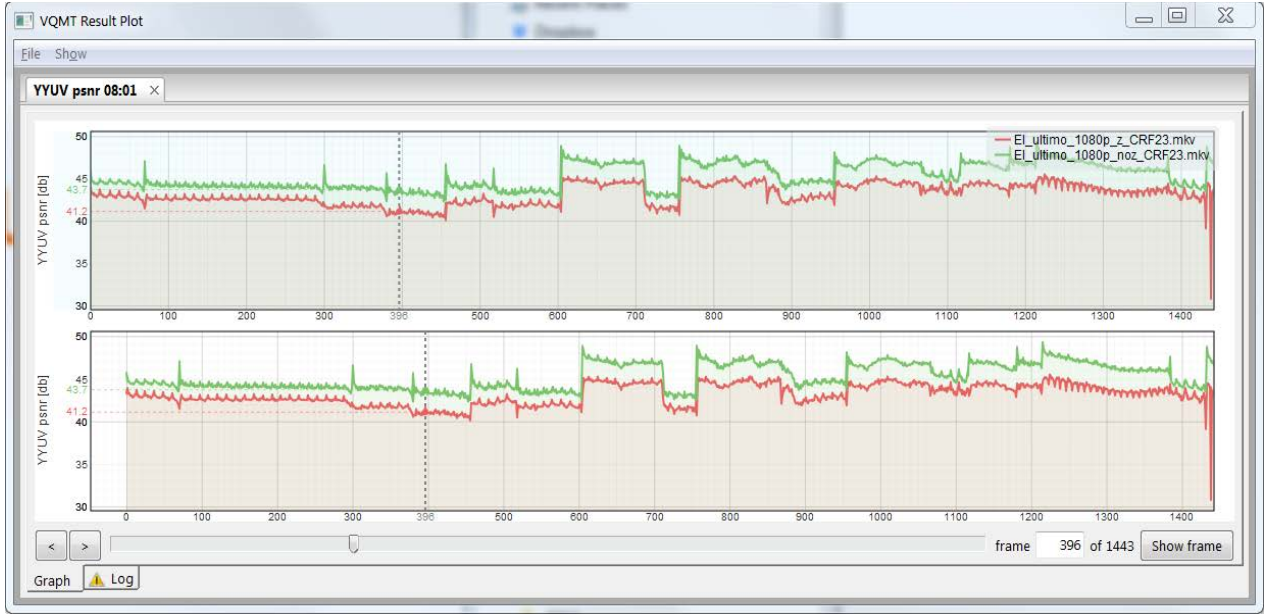

00:00:16:15
ZPEG
00:00:16:15
NO ZPEG

El-Ultimo

PSNR scan: no major qualitative differences;

ZPEG processed file comes out of alignment at very end accounting for red spike on the extreme right (this happened frequently, but only at file end).

Top/Bottom analysis: Slight loss of vibrance; see video.



Ironman

Description: Ironman is a simple 2D animated movie.

Data rate reduction: 3.27%
VMAF differential: -5.43
PSNR differential: -3.18

Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.


ZIP
NO ZIP

Ironman

PSNR scan: no major qualitative differences;

ZPEG processed file comes out of alignment at very end accounting for red spike on the extreme right.

Top/Bottom analysis: Nothing significant noted.

Sintel

Description: Sintel is a complex animated movie with lots of motion and fine detail. Image shown on the right shown in full screen on the following slide.

Data rate reduction: -10.17%
VMAF differential: -6.13
PSNR differential: -2.43

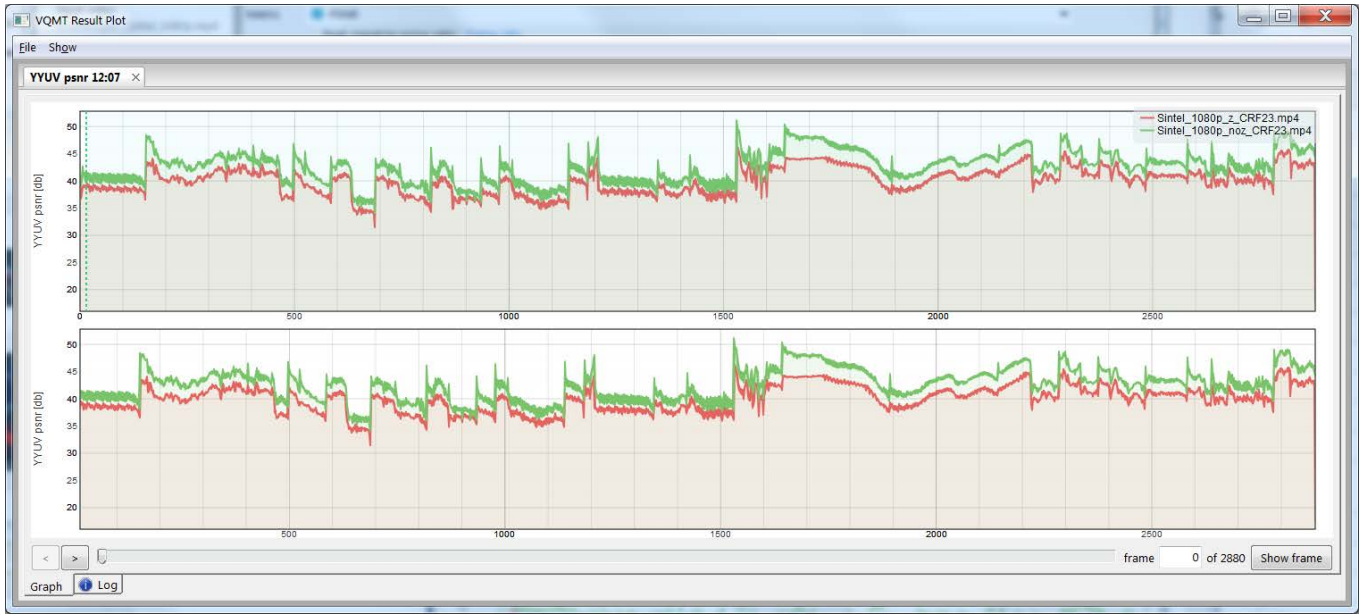
Subjective quality analysis: No visual difference in real-time or frame-by-frame playback. Excellent retention of detail.


ZIPEG
NOZIPEG

Sintel

PSNR scan: no major qualitative differences;

Top/Bottom analysis: No major differences noted.



Spongebob

Description: Spongebob is a trailer for the Spongebob Squarepants movie

Data rate reduction: -10.17%
VMAF differential: -3.44
PSNR differential: -1.33

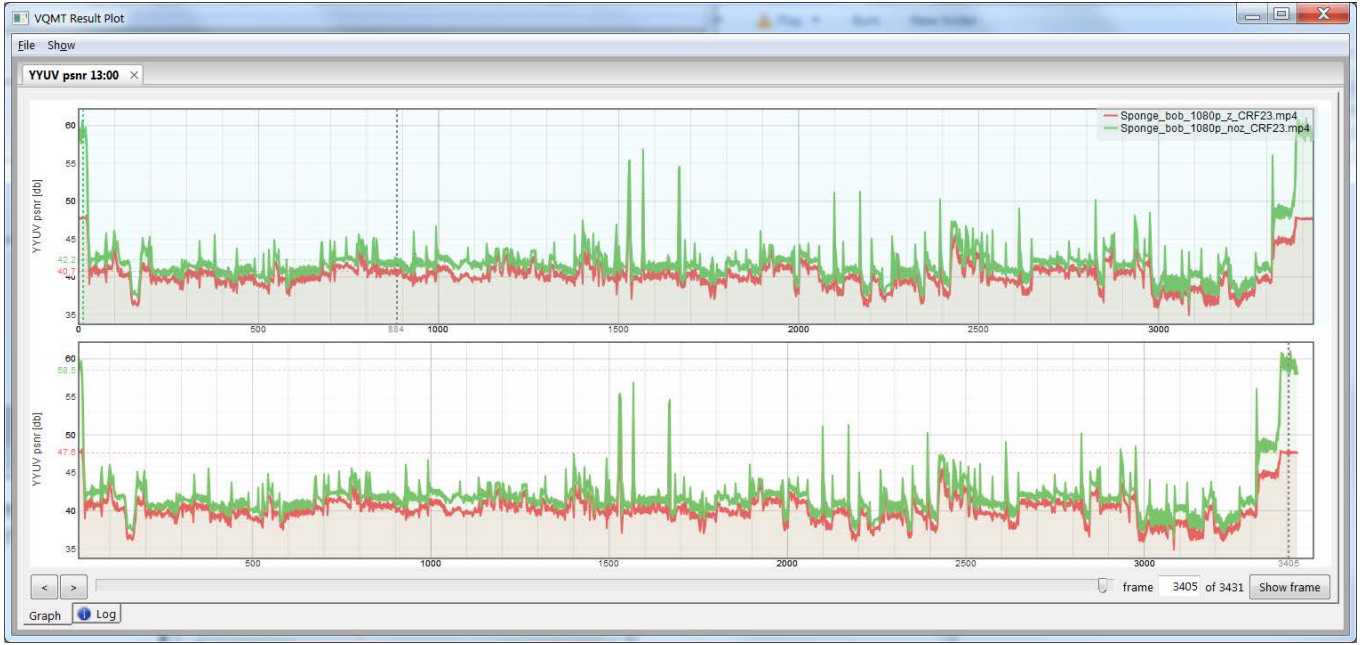
Subjective quality analysis: No visual difference in real-time or frame-by-frame playback.


00:00:20:20
ZPEG
00:00:20:20
No ZPEG

Spongebob

PSNR scan: Checked differential at end of clip; no subjective difference (at very high end of PSNR scale so no visual differences anticipated).

Top/Bottom analysis: No major differences noted.

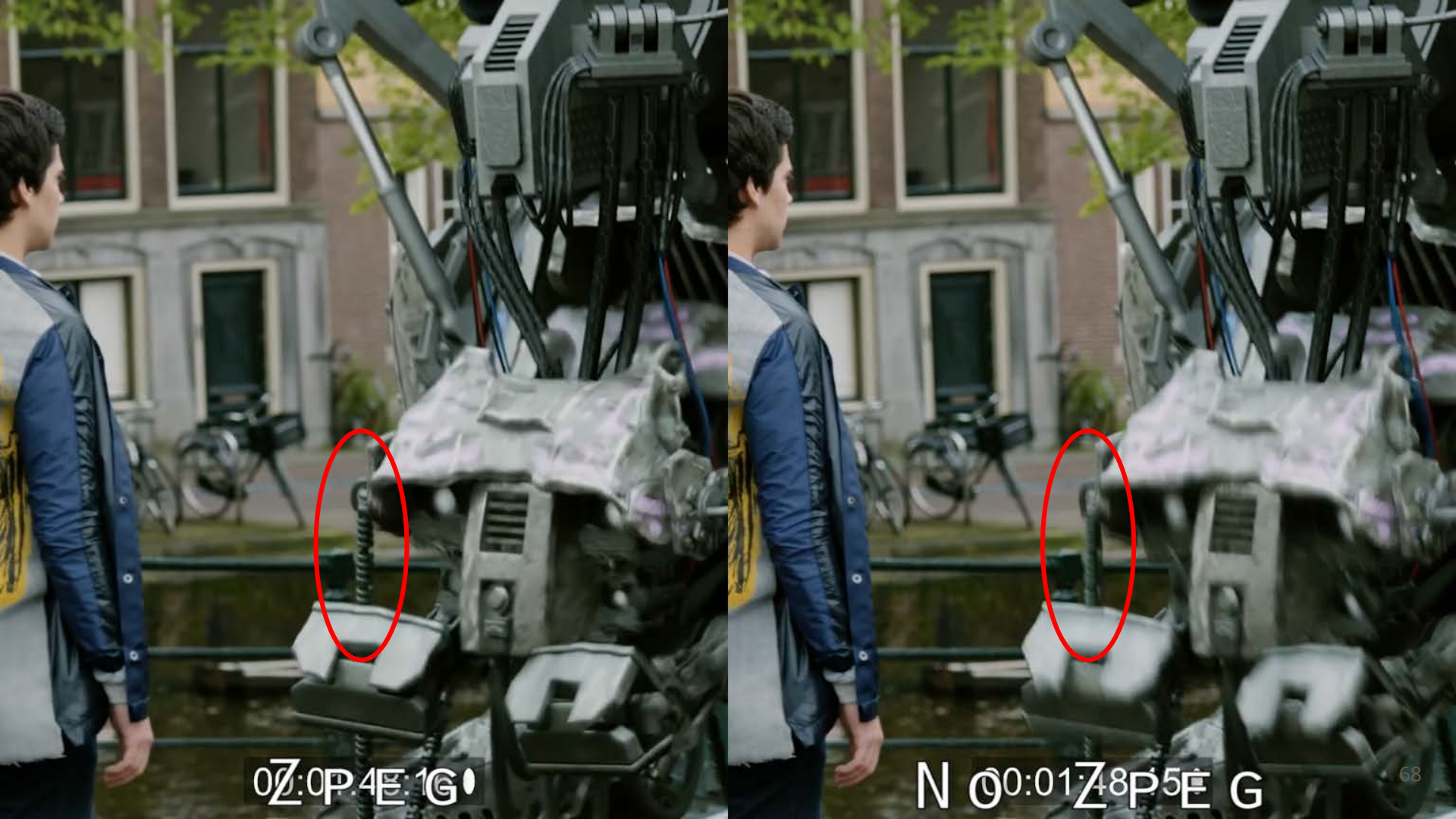


Tears of Steel

Description: Tears of Steel is a clip from the Blendor movie; part real world, part CG.

Data rate reduction: -10.43%
VMAF differential: -5.99
PSNR differential: -2.18

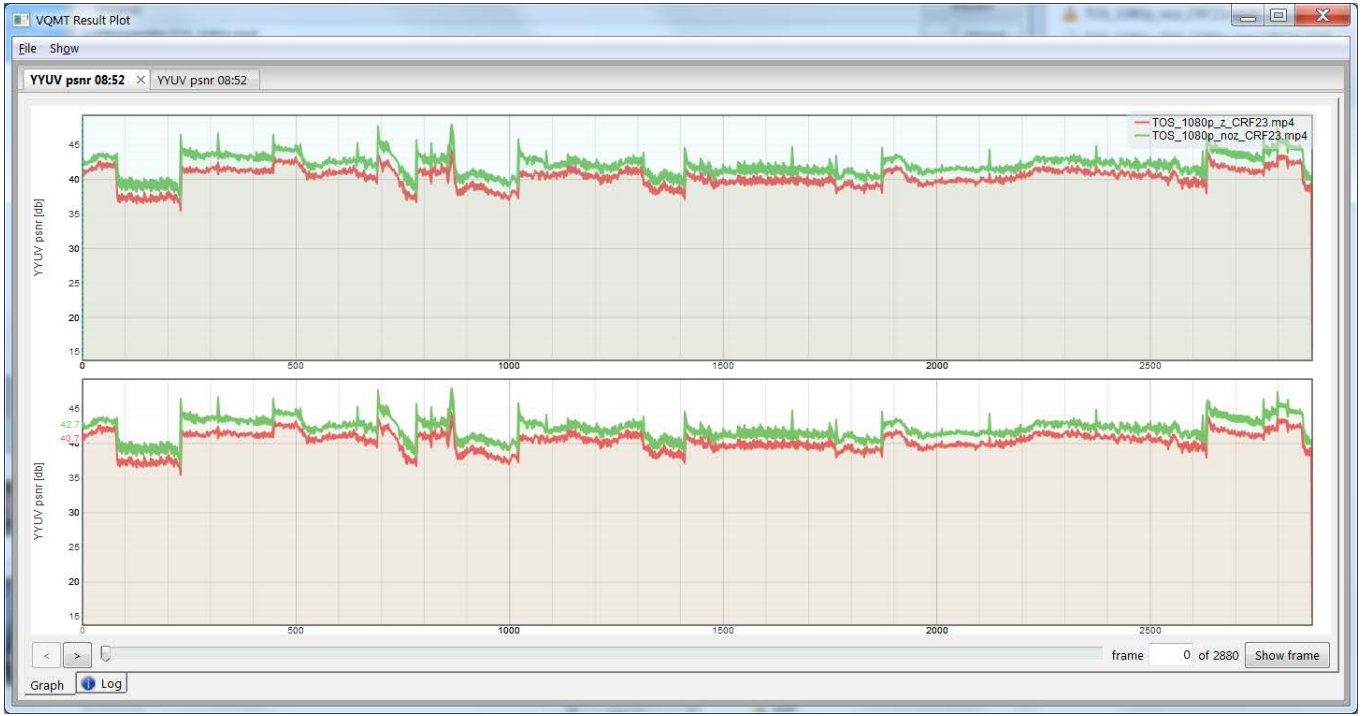
Subjective quality analysis: No visual difference in real-time playback. Slight loss of detail in frame-by-frame playback; would not be noticed by real-time viewers.


ZOOM
NOZOOM

Tears of Steel

PSNR scan: no major qualitative differences;

Top/Bottom analysis: ZPEG removes film grain and some minor detail. Would not be noticed by real-time viewers. See video.



Summary

Movie-ish content - very good performance here with a 27.54% data rate reduction **on top of** CRF. The only cautions here relates to reduction in film grain.

Business-content - Not indicated for screencams or PowerPoint content, but excellent performance with other videos, averaging over 31% data rate reduction with no artifacting.

Animated content - not a particular strength of ZPEG, but did no harm in any clips and produced an overall 11.52% savings.

Overall Findings:

1. The ZPEG Engine for x264 bitrate savings averaged 21.25% over all test clips.

2. ZPEG Engine processing never introduced any noticeable artifacts in any tested videos, despite data rate reductions in excess of 38% in some cases.

    a. Early testing with a beta version of the ZPEG Engine detected visible color shifts in A/B testing. This problem was completely fixed with the latest tested version.

    b. Later testing detected out of order frames at the end of the PSNR sequence. This problem was completely fixed with the latest tested version.

3. ZPEG Engine processing would not be noticeable by typical viewers in real-time playback.