APPLYING QUALITY ENCODING METRICS FOR 2D AND VR VIDEO

Jan Ozer
www.streaminglearningcenter.com
jozer@mindspring.com/
276-235-8542
@janozer

Agenda

- Overview of Objective Quality Metrics
- Tools/Metrics I use
- What about VR?
- Finding the floor and ceiling
  - 2D
  - VR
- Building your encoding ladder
- Other configuration options

What Are Objective Quality Metrics

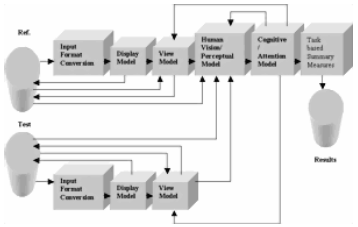
- Mathematical formulas that (attempt to) predict how human eyes would rate the videos
  - Faster and less expensive
  - Automatable
- Examples
  - Peak Signal to Noise Ratio (PSNR)
  - Structural Similarity Index (SSIM)
  - SSIMPlus
  - VMAF
  - Various VR metrics

Subjective vs. Objective Visual Quality

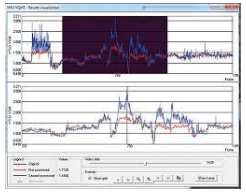
Standards-based

Informal

Perceptual Quality Analyzers



Mathematical (MSE-based)



What are they?	Formal standards	Informal	Perceptual Quality Analyzers	Pure Math-based Quality Models
Example	ITU-T P.910 recommendation	Golden Eye Testing	PQA (Tek), DMOS, SSIMplus, VMAF (Netflix)	PSNR, SSIM
Pros	Gold standard	Accessible	Fast, simple to apply, good correlation to subjective	Fast, simple to apply, cheap
Cons	Time consuming, inappropriate for production	Time consuming	Expensive Some are proprietary	Low correlation with subjective benchmarks

Measure of Quality Metric

- Role of objective metrics is to predict subjective scores
- Correlation with Human MOS (mean opinion score)
  - Perfect score - objective MOS matched actual subjective tests

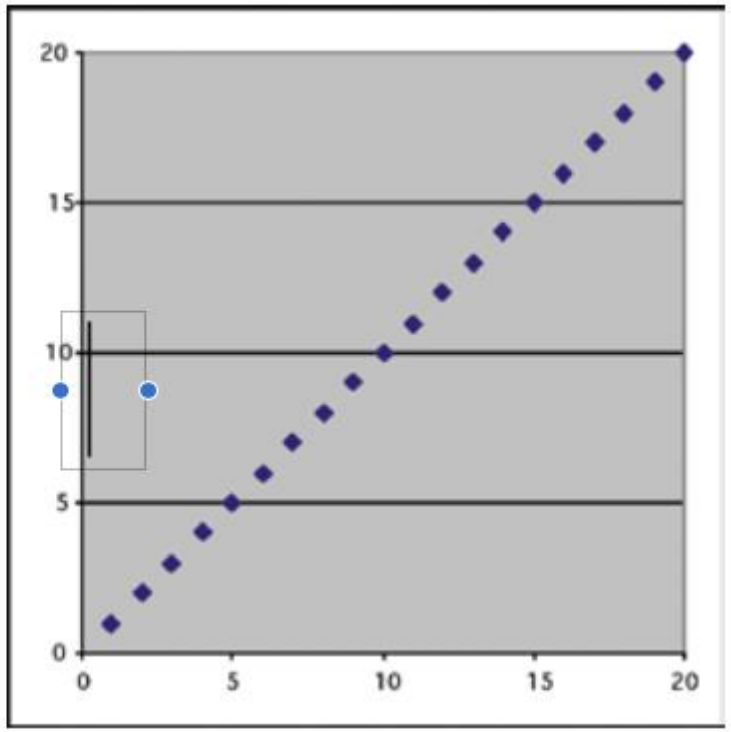
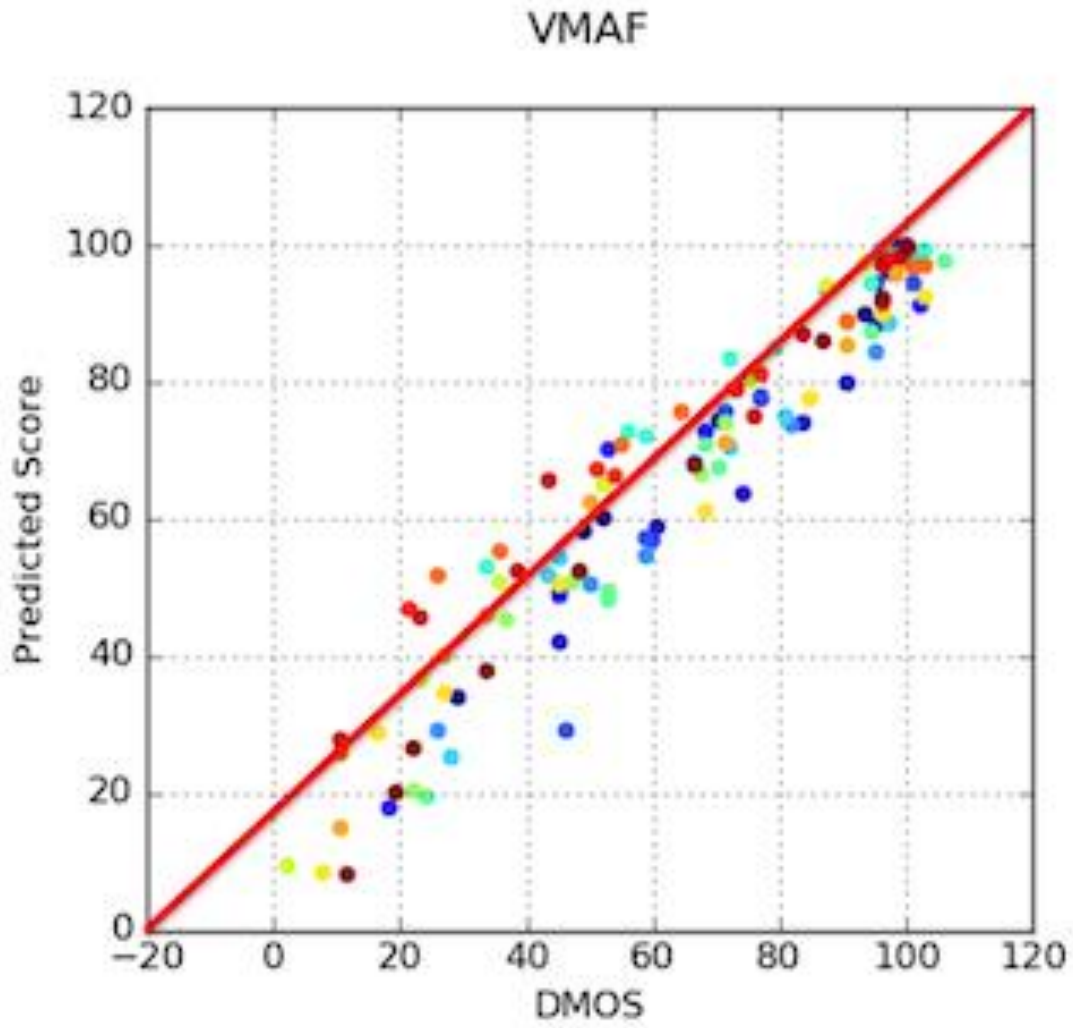
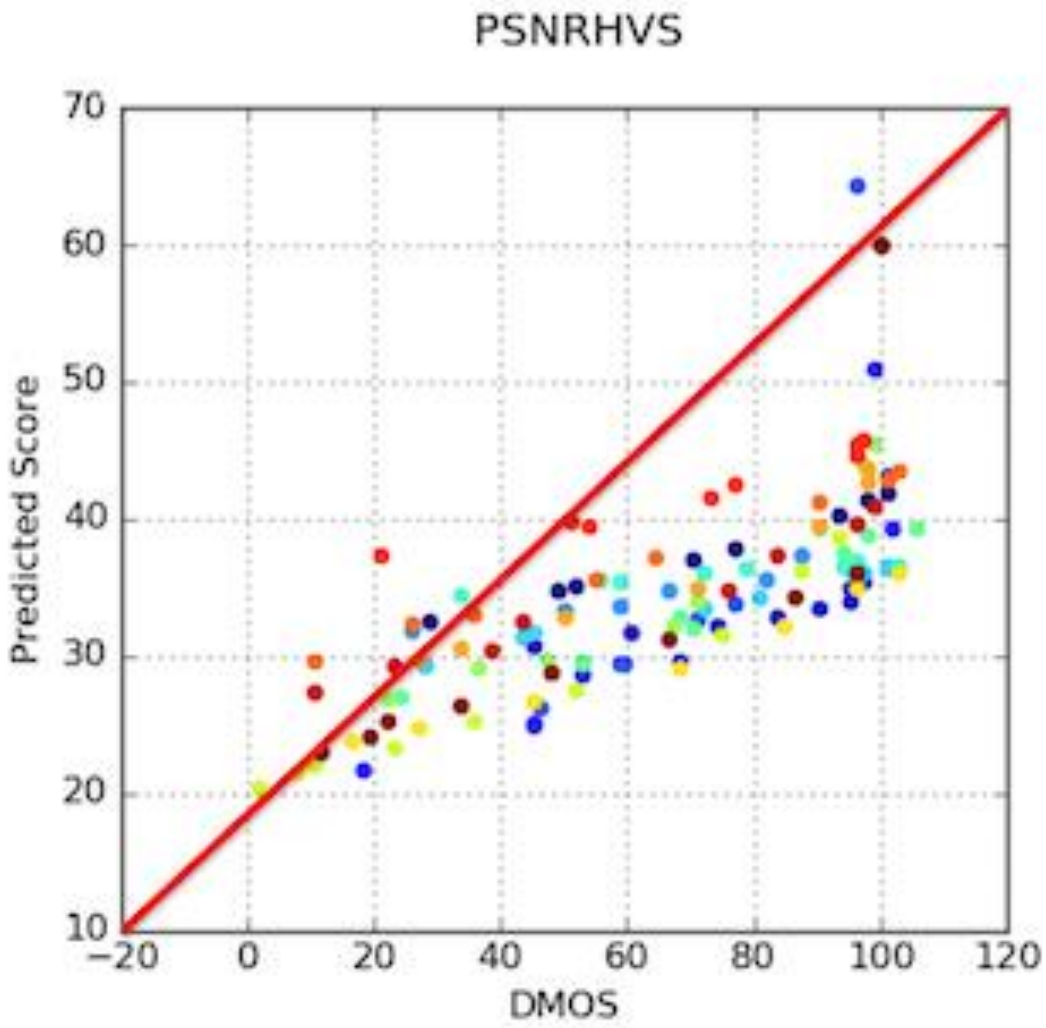


Figure 10. Correlation coefficient: 1.

Measure of Quality Metric

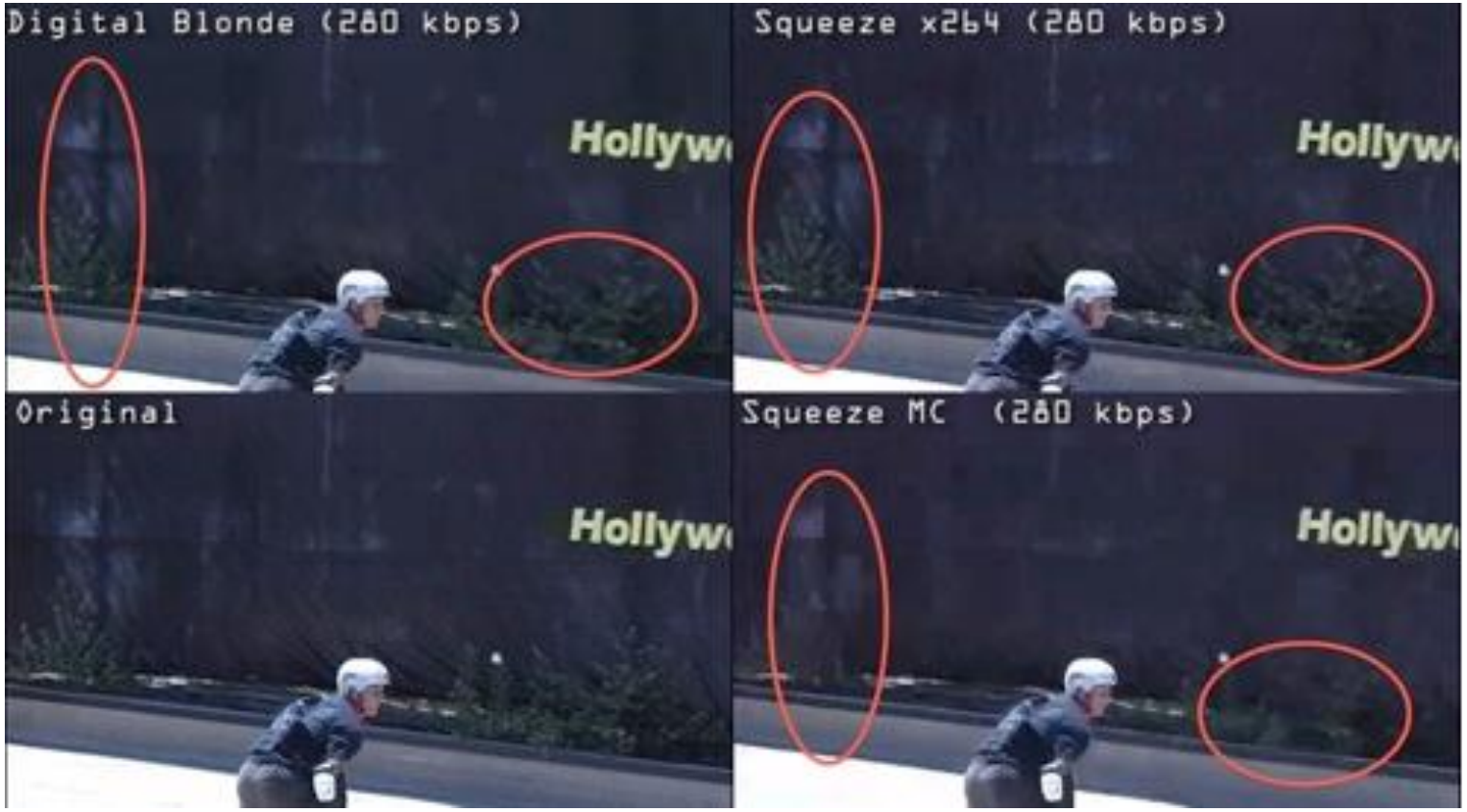




The Bottom Line

- Every new metric presents similar accuracy analyis
- You have to use the metric, confirm with subjective comparisons, and get comfortable with a rating

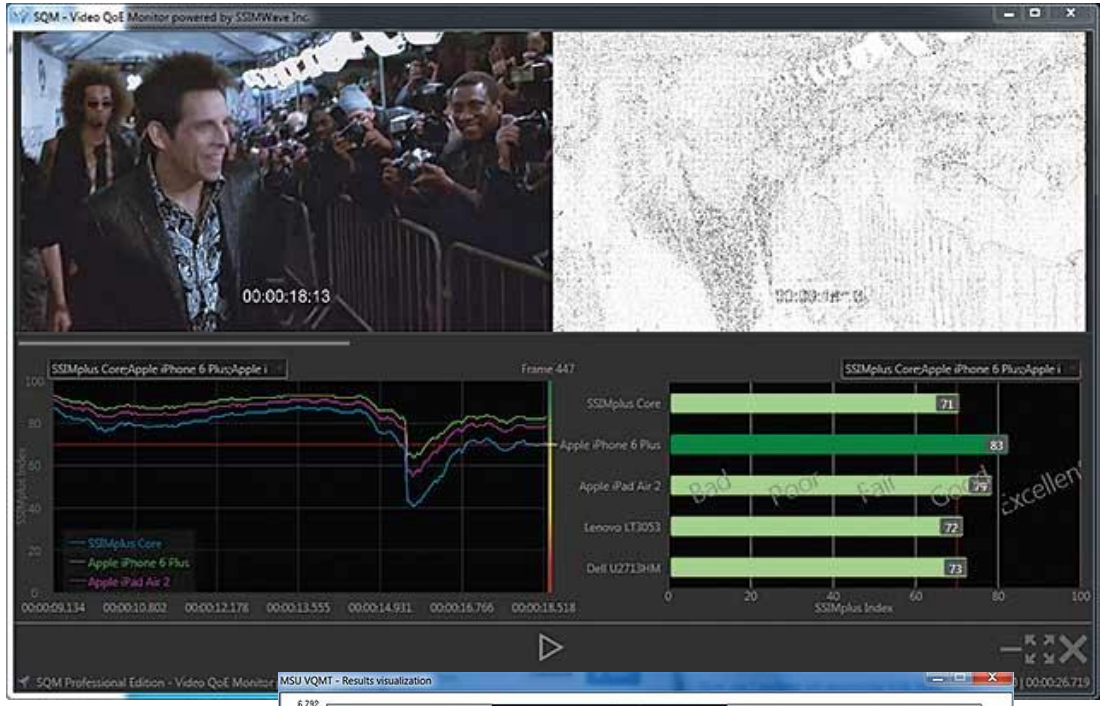
Took Me From Here

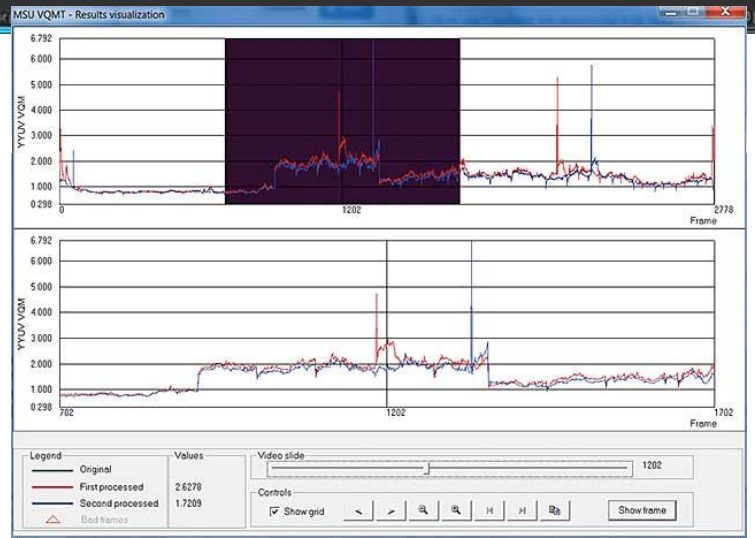


Time consuming and error prone
Subjective comparisons

To Here

VQM (lower is better)					
	Codec A	Codec B	Codec C	High > Low	Codec A > Codec B
Office 1	0.36	0.36	0.37	-3.54%	0.61%
Office 2	0.69	0.61	0.70	-13.51%	12.32%
Office 3	0.28	0.28	0.32	-14.74%	1.32%
Office 4	0.87	0.79	0.87	-9.63%	9.63%
Parking 1	0.68	0.61	0.74	-21.23%	10.90%
Parking 2	0.57	0.55	0.64	-15.47%	3.04%
Parking 3	1.86	1.58	1.76	-17.88%	17.88%
Parking 4	0.47	0.49	0.51	-8.86%	-3.81%
Retail 1	0.56	0.54	0.56	-4.27%	4.27%
Retail 2	0.68	0.66	0.69	-4.45%	3.39%
Retail 3	0.78	0.72	0.76	-8.64%	8.64%
Retail 4	0.73	0.67	0.88	-32.16%	8.52%
Traffic 1	0.55	0.50	0.58	-15.89%	9.14%
Traffic 2	0.34	0.32	0.38	-17.79%	6.39%
Traffic 3	0.52	0.49	0.55	-11.42%	5.29%
Traffic 4	0.68	0.61	0.66	-11.56%	11.56%
Total	**10.61**	**9.78**	**10.96**		
7.84%	Difference between Codec A and Codec B				
-3.34%	Difference between Codec A and Codec C				
-12.13%	Difference between Codec B and Codec C				
	0.61				
	Green equals best in category				
	Orange means worst in category				
	Difference greater than 7.5%				





Statistically meaningful comparisons

With Objective Quality Metrics You Get

- More data
  - Can run many more tests in much less time
- Better data
  - Mathematical models can measure smaller changes than your eye can easily discern
- High level operation
  - Input source and test file(s)
  - Test program delivers a score

Trust, But Verify

- Never rely solely on objective test results
- Compare files yourself to verify comparisons
  - Still image comparisons
  - Side by side real time playback

The Tools I use

- Moscow State University Visual Quality Comparison Tool (VQMT)
- Hybrik Cloud Encoding Analysis Tool
- SSIMWave Video Quality-of-Experience Monitor (SQM)
  - From one of the inventors of SSIM metric

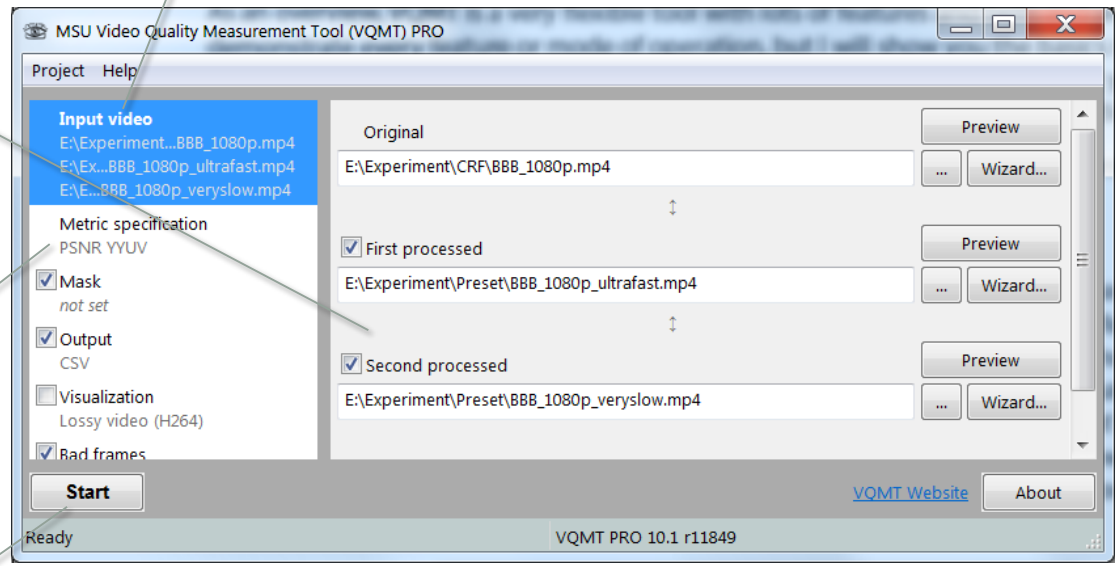
VQMT Workflow

Load Source File

Load one or two encoded files

Choose Metric

Press Start



Results Visualization

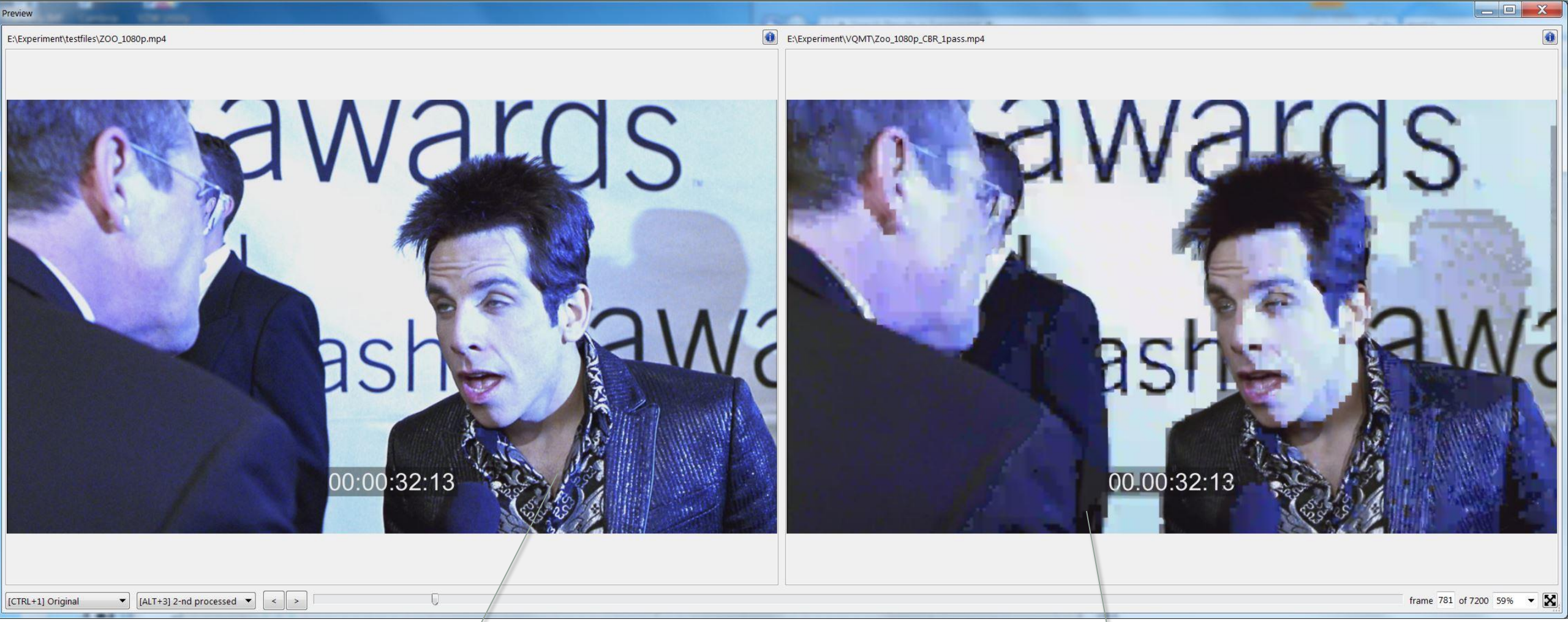
Score entire comparison

Orange– first file
Green– second

Zoom in of selection

Slide through frames

Click to Show Actual Frames



VBR-encoded

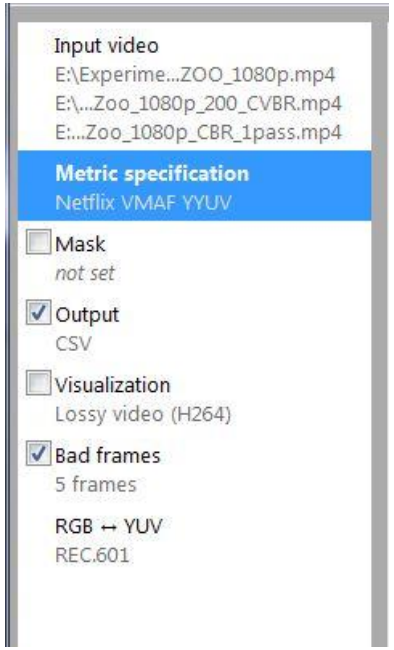
- Slide through video file
- Compare

CBR-encoded

- Compare side-by-side or hot key between original and two encoded files

Bad Frames Feature

- It’s not always about overall score
  - Preview feature lets you scan through frames manually
  - Bad frames saves lowest quality frames



- Overall delta between CBR and VBR is minor (90.83 vs 90.27)
- Lowest quality frames ID transient issues that would lower viewer QoE

MSU VQMT

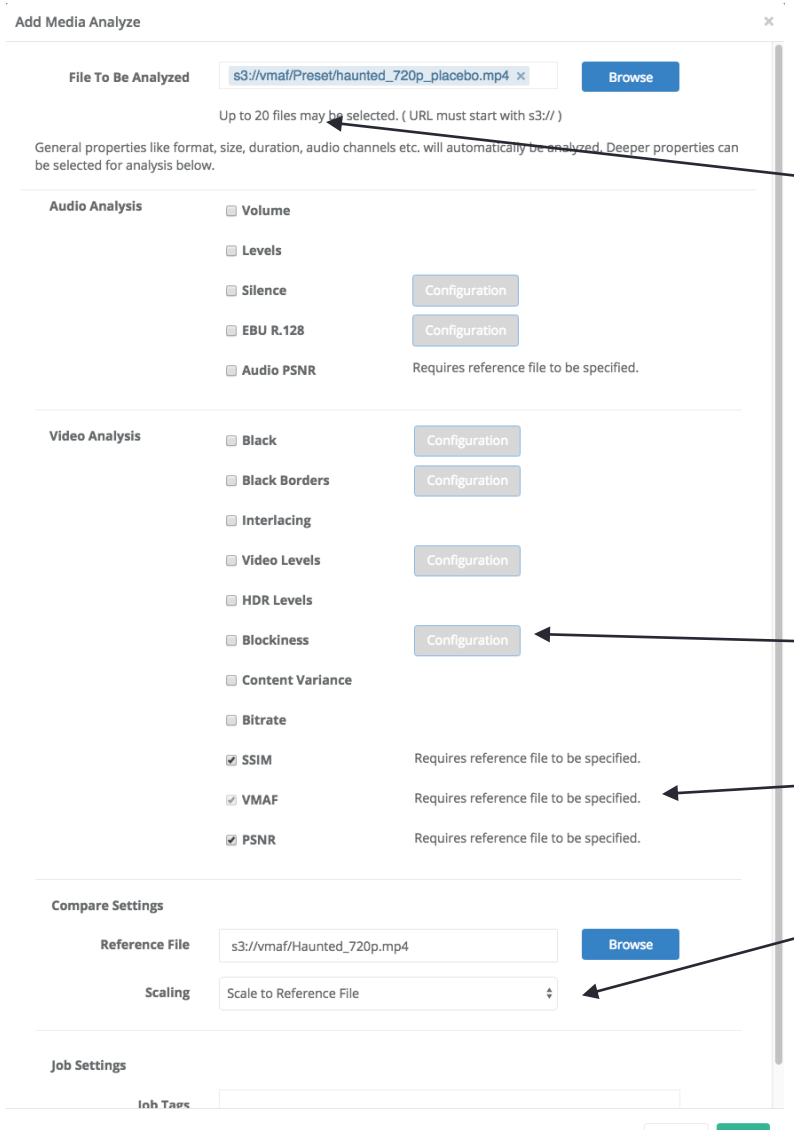
Pros

- Affordable (~$995)
- GUI and command line
- Very visual – easy to see test results in actual frames
- Multiple algorithms – VMAF, PSNR, SSIM, MS SSIM
- My review of VQMT
  - bit.ly/VQMT_review

Cons

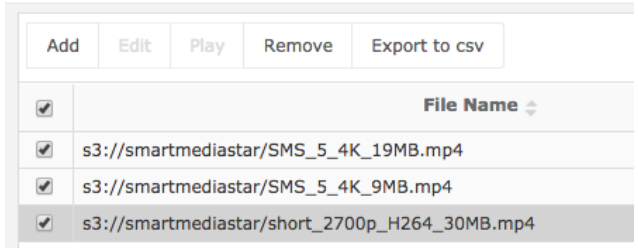
- Can only compare files of:
  - Like resolution
  - Like frame rate
  - Are workarounds (scale to YUV), but cumbersome
- Time consuming data entry

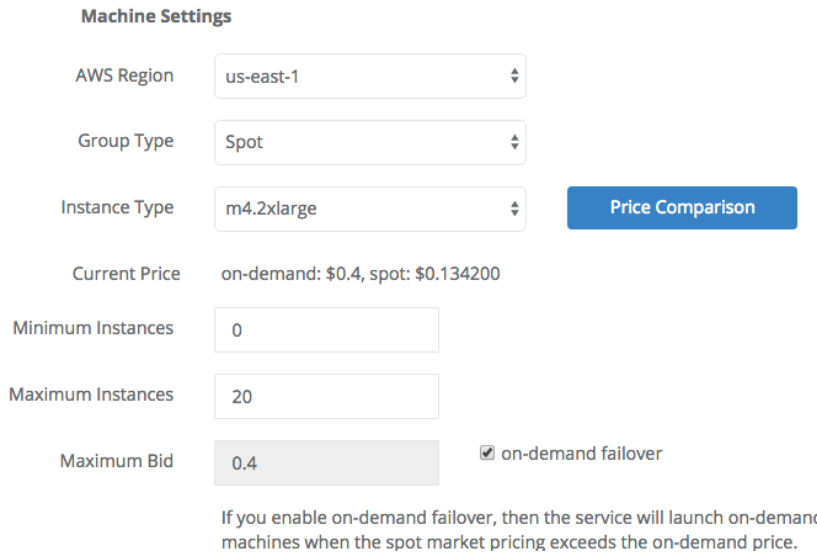
Hybrik Cloud Media Analyzer



- Industrial Strength Metrics
  - Compare up to 20 files at once (save input time)
  - Many file related metrics
- One no-reference metric (blockiness)
- Three full-reference metrics
  - SSIM
  - VMAF
  - PSNR
- Automatically scales to reference file size

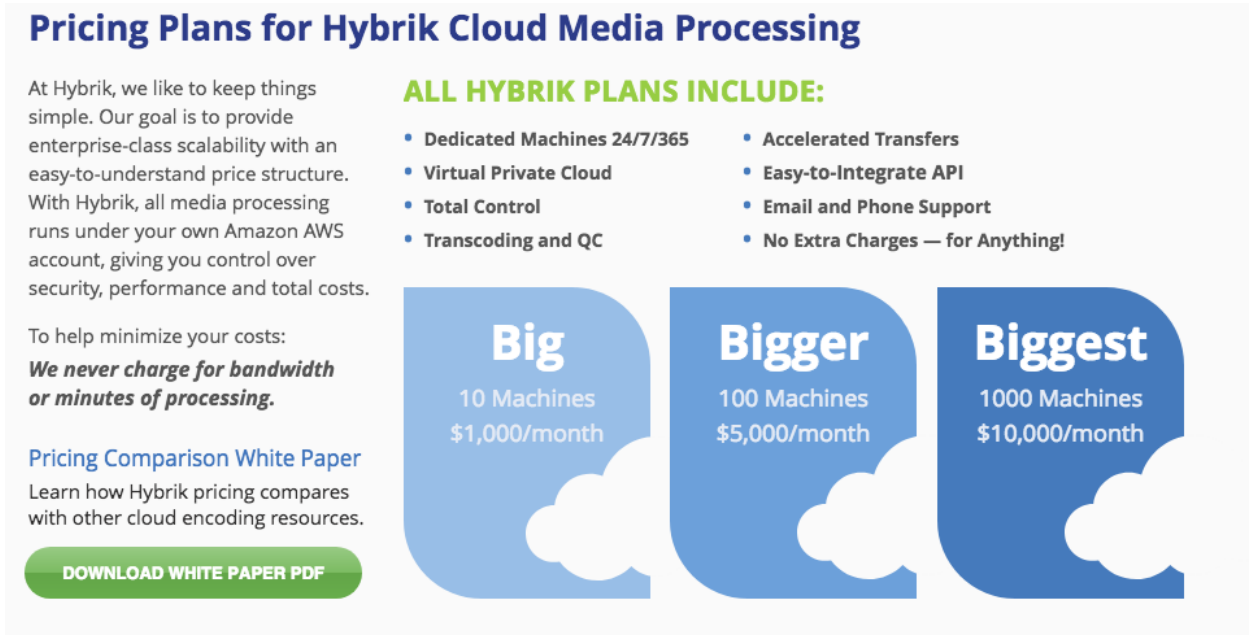
Hybrik Cloud Media Analyzer





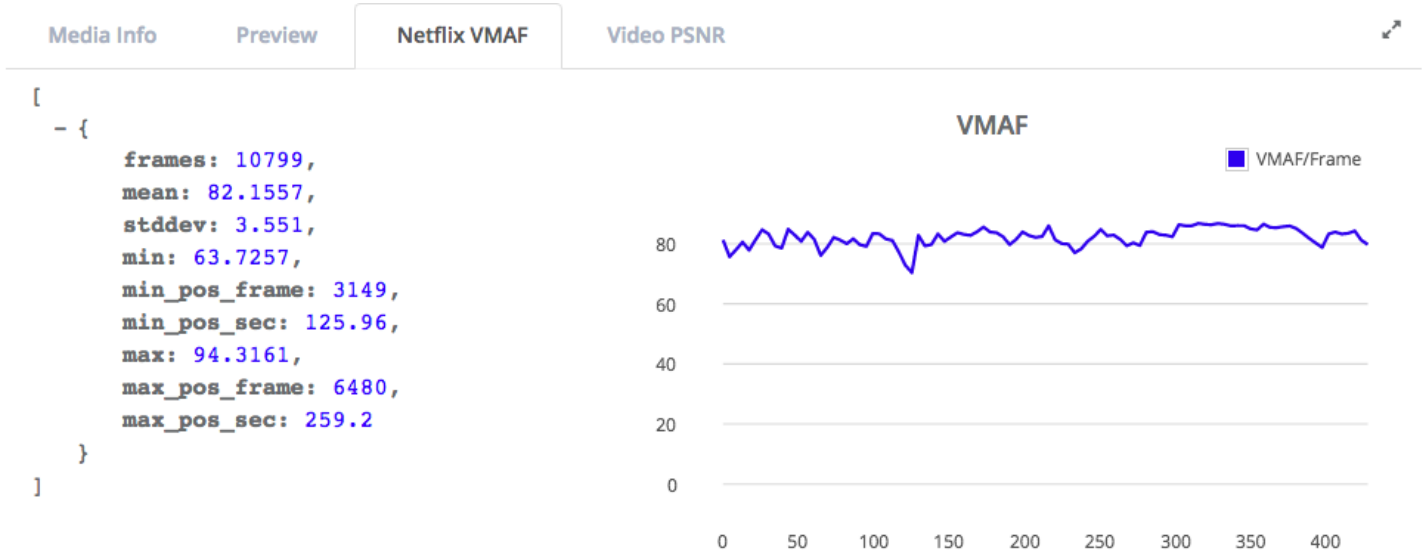
- Export multiple results to CSV
  - Import directly to spreadsheet
  - Saves lots of time-consuming and boring input time
- Scalable, so very fast
  - Single VMAF analysis with MSU can take 10 – 15 minutes
  - Can scale somewhat on my 40-core HP Z84, but still limited
  - With Hybrik, can scale up to full capacity
- **Can reduce human/machine time on some projects from multiple hours to minutes**

Hybrik Model



- Full service cloud encoder; no analyis-only pricing (despite my best efforts)
- Bottom line: When you need throughput, you need to harness the cloud
- Caveat: I have done consulting work for Hybrik; I'm not being compensated for this mention

Hybrik Limitations



- Limited visualization; no frame viewing (coming)
- As with all cloud services, can be start-up latency
  - Not great for onesies and twosies

Hybrik

- Here at NAB – Booth SU9906CM

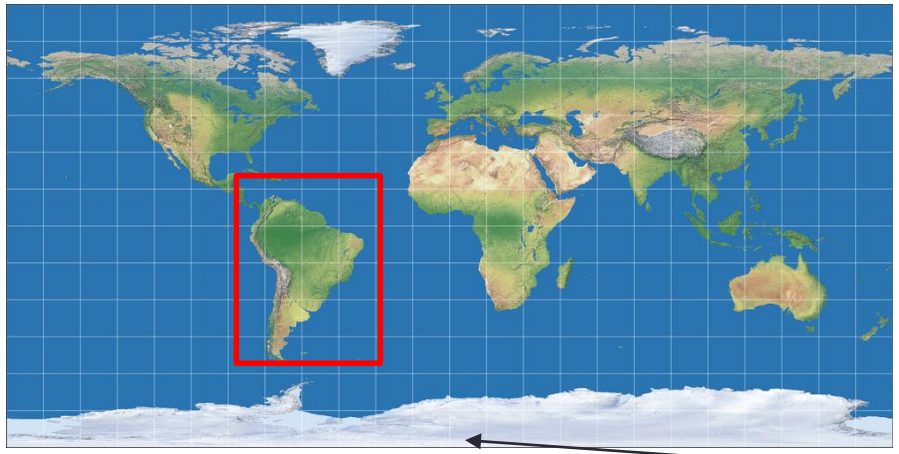
SQM Overview

- Based on SSIMplus Algorithm
- Rates videos on scale that corresponds with human perception
  - 80 – 100 – Excellent
  - 60 – 80 – Good
  - 40 – 60 – Fair
  - 20 – 40 – Poor
  - < 20 – Bad
- Predicts ratings on multiple devices
  - Phones, TVs, monitors, etc.
- Separate command line tool for Windows/ Linux
- My review
  - http://bit.ly/SQM_review
- Here at NAB - South Upper Hall booth #5224.

What about VR?

- The problem
- Solutions
- The workaround

The Problem



- Multiple VR storage formats
  - Equirectangular above is most common
    - Heavily distorted at poles
  - All represent 360 image in flat world
- VR is 360
  - Relatively similar in the middle
  - Heavily distorted at poles

Issues

- General
  - Where is viewer looking?
    - Is this relevant?
    - Can we weight by presumed focus of attention?
    - Should we?

- General
  - Do flat metrics work?
    - If so, which?
  - What VR metrics are available?
    - Do they work?

Tools and Metrics

- There are multiple VR metrics
  - They are not generally accessible
  - None in MSU, SQM, or Hybrik

Reviews are Mixed

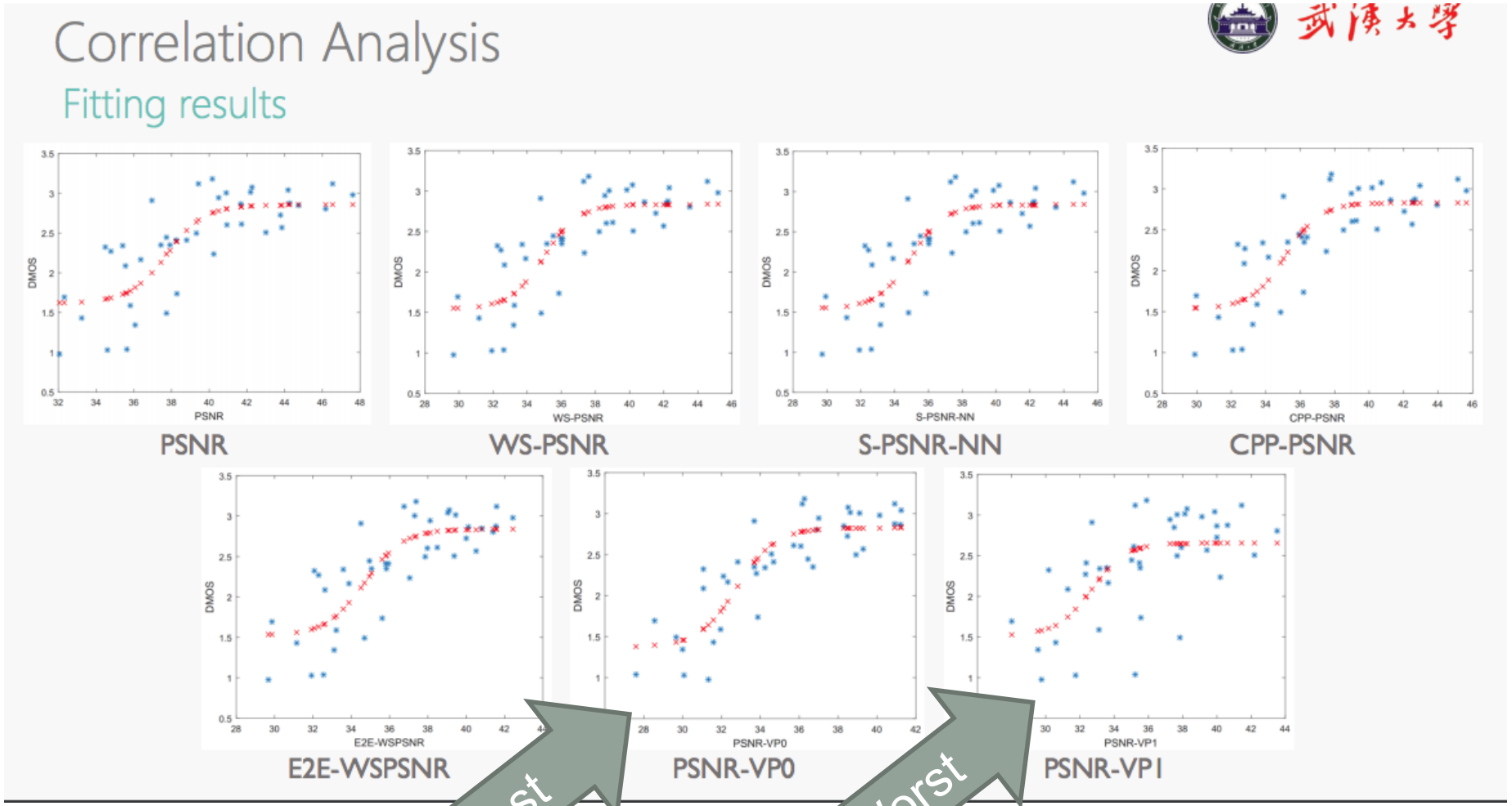
- On the Performance of Objective Metrics for Omnidirectional Visual Content (http://bit.ly/vrqm_1), "Objective metrics specifically designed for 360-degree content ***do not outperform conventional methods*** designed for 2D images."
- An evaluation of quality metrics for 360 videos (http://bit.ly/vrqm_2), "Most objective quality measures are well correlated with subjective quality. Among the evaluated quality measures***, [traditional flat] PSNR is the most appropriate for 360 video communications."***
- Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video (http://bit.ly/vrqm_3), "Our method makes the quality evaluation results ***more accurate and reliable since*** it avoids error propagation caused by the conversion from resampling representation space to observation space."

Benchmarking Virtual Reality Video Quality Assessement (http://bit.ly/vrqm_4)

Objective model	Description
PSNR	Peak Signal-to-Noise Ratio. Calculates PSNR based on all samples with equal weight.
WS-PSNR	Weighted to Spherically uniform PSNR. Calculates PSNR based on all samples with a weighted parameter, depending on the sample area on the spherical surface.
S-PSNR-NN	Spherical PSNR without interpolation. Calculates PSNR based on a point set evenly sampled on the sphere surface, whose value is taken from the nearest neighbor integer sample positions to avoid the impact due to interpolation filters.
CPP-PSNR	PSNR for Carster Parabolic Projection. Compares quality across different projection methods using Carster Parabolic Projection format.
E2E-WSPSNR	End to End WS-PSNR. Proposes end to end assessment for comparing compression performance of different projection.
PSNR-VP0 and PSNR-VP1	Calculates PSNR on 2D displays with the two viewports (VPs) rendered from the decoded bit stream with predefined parameters.

- Evaluated these metrics

Benchmarking Virtual Reality Video Quality Assessement (http://bit.ly/vrqm_4)

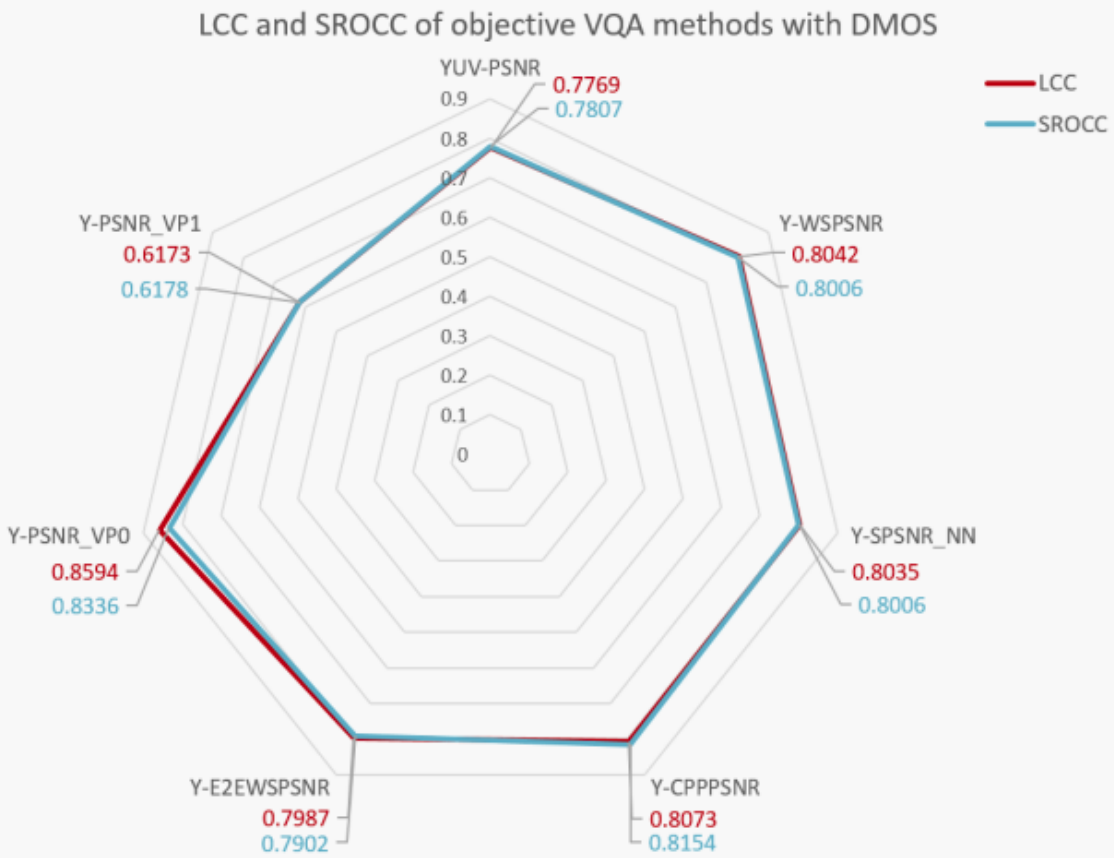


Benchmarking Virtual Reality Video Quality Assessement (http://bit.ly/vrqm_4)

Not night and day difference

Higher is Better

Algorithm	LCC	SROCC
PSNR	0.7769	0.7807
WS-PSNR	0.8042	0.8006
S-PSNR-NN	0.8035	0.8006
CPP-PSNR	0.8073	0.8154
E2E-WSPSNR	0.7987	0.7902
PSNR-VP0	**0.8594**	**0.8336**
PSNR-VP1	0.6173	0.6178



What I've Done

- All work performed for Pixvana; ***data courtesy Pixvana***
  - Compared Samsung WS-PSNR with PSNR and VMAF
- Focus
  - Utility for choosing appropriate data rate for switching resolutions in ABR ladder
    - Less convenient than PSNR/VMAF
    - Is it worth the effort

- Compared Samsung WS-PSNR with PSNR and VMAF
  - https://github.com/Samsung/360tools

Supported formats

```
+       Equirectangular projection (ERP)
+       Icosahedral projection (ISP)
+       Octahedron projection (OHP)
+       Cubemap projection (CMP)
+       Truncated Square Pyramid projection (TSP)
+       Segmented Sphere Projection (SSP)
+       Reshaped Icosahedral projection (RISP)
+       Reshaped Octahedron projection (ROHP)
+       Reshaped Cubemap projection (RCMP)
```

Supported quality metrics

```
+       PSNR - conventional Peak Signal to Noise Ratio quality metrics
+       S-PSNR - spherical PSNR (requires sphere_655362.txt file with point coordinates)
+       WS-PSNR - weighted Spherical PSNR (for equirectangular projection only)
+       CPP-PSNR - equal area common projection PSNR
```

Building Encoding Ladder

- Netflix-like method
  - Top rate determined by budget or minimum quality
  - Lower data rates distributed by formula (so rungs between 1.5/2x apart)
  - Use quality metric to choose resolution at each rate
  - Did WS-PSNR provide substantially different result than PSNR

Zap1 - VMAF	4K	2K	1080p	720p	480p	360p	240p
5000	90.19	89.70	84.82				
4500	89.58	88.23	84.38				
4000	88.43	87.50	83.84				
3800	87.88	87.14	83.58				
3600	87.27	86.71	83.25				
3400	86.60	85.72	82.87				
3200	85.80	85.40	82.45				
3000	85.03	85.09	82.01				
2800	83.97	84.34	81.43				
2600	82.86	83.50	80.85				
2400	81.45	82.51	80.09	71.92			
2200	79.79	81.24	79.20	71.35			
2000	77.94	79.82	78.04	70.66			
1800		78.11	76.73	69.70	53.28		
1600		75.91	74.93	68.41	52.82		
1400		73.26	72.64	66.89	52.13	32.07	
1200		69.83	69.69	64.68	51.05	31.75	
1000		65.15	65.75	61.64	49.36	31.17	
900		62.26	63.25	59.64	48.11	30.76	
800		58.69	60.27	57.20	46.54		
700		54.29	56.62	54.06	44.63	29.18	
600		48.79	52.32	50.65	42.02	27.84	
500			46.65	45.96	38.74	25.92	
400			39.06	40.23	34.21	23.11	
300			28.52	32.68	x		
200			13.88	21.73	x		

Building Encoding Ladder

	Video 1			Video 2			Video 3		
	VMAF	PSNR	WS-PSNR	VMAF	PSNR	WS-PSNR	VMAF	PSNR	WS-PSNR
4K > 2K	8,000	5,000	5,000	3,000	2,200	2,000	8,000	5,000	5,000
2K > 1080p	3,200	2,000	1,800	1,000	900	900	3,200	2,000	1,800
1080p > 720p	1,000	1,000	1,000	400	500	400	1,000	1,000	1,000
720p > 480p	NA	500	500	NA	100	100	NA	500	500
480p > 360p	NA	300	200	NA	NA	NA	NA	300	200

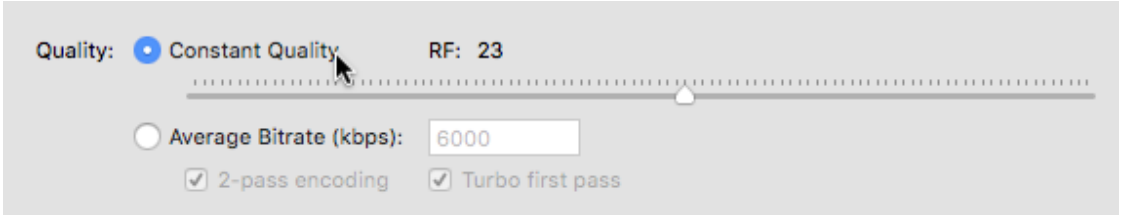
- Not really
- Three different files
  - Switch points very different between VMAF and PSNR/WS-PSNR
- On these three files, however, PSNR/WS-PSNR deliver about the same result
- Conclusion: PSNR/VMAF both more accessible, faster, so WS-PSNR adds no value in this application

Using Quality Metrics – Finding the Floor

- About CRF
- Configuring ladder
  - Floor and ceiling
  - 2D/3D

About Constant Rate Factor Encoding

- Encoding mode available in x264, x265, VP8/9
- Encodes to a specific *quality* level, not a data rate
- Two uses
  - As gauge of encoding complexity
  - With caps, a per-title encoding technique



- Range is 1-51
  - Lower number means higher quality
  - For 2D video, CRF 23 roughly delivers Hollywood (iTunes) quality

Finding the Optimal Data Rate for 2D Content (Per-title)

- Compute data rate with CRF 23
  - Values varied from 1,001 to 6,111 (over 600%)
- Measure VMAF rating
  - Values ranged from 92.74 to 96.88
  - Standard deviation was 1.39 (pretty small)
- Analysis
  - At 2.7 Mbps, a talking head video offers same quality as movie at 6.1 Mbps (even more for synthetic videos)
  - Validating the benefits of per-title encoding

CRF23 - 1080p	FPS	Description	Data Rate	VMAF
Tears of Steel	24	Real world/CG movie	4,747	96.45
Sintel	24	Complex animation	5,168	96.96
Big Buck Bunny	30	Simple animation	3,657	96.88
Screencam	30	Camtasia-based video	1,625	96.59
Tutorial	30	PowerPoint and talking head	1,001	96.68
Talking Head	30	Simple talking head	2,706	95.47
Freedom	30	Concert footage	5,527	95.90
Haunted	30	DSLR movie-like production	6,111	92.74
Average			**3,818**	**95.96**
Standard deviation				**1.39**

- Conclusion:
  - CRF maps accurately to VMAF values

Hollywood Verification

Program	Owner	Width	Height	Frame Rate	Target Video DR	Bits Per Pixel
Angie Tribeca	Tuner	1916	1076	23.976	5,060	0.102
Better Call Saul	Sony	1916	1076	23.976	5,169	0.105
Blackish	ABC/Disney	1920	1080	23.976	4,953	0.1
Brooklyn 999	Universal	1920	1080	23.976	5,094	0.102
Family Guy	FoxTV	1920	1080	23.976	5,173	0.104
Fresh of the Boat	20th Century	1920	1080	23.976	4,946	0.099
Full Frontal	TBS	1920	1080	23.976	5,238	0.105
I am Cait	E!	1440	1080	23.976	5,261	0.141
Sherlock	BBC	1920	1080	25	5,062	0.098
The Affair	Showtime	1912	1080	23.976	4,959	0.1
Last Man on Earth	20th Century	1920	1080	23.976	5,117	0.103
Transformers	Hasbro	1920	1080	23.976	5,128	0.103
Average					5,097	0.105

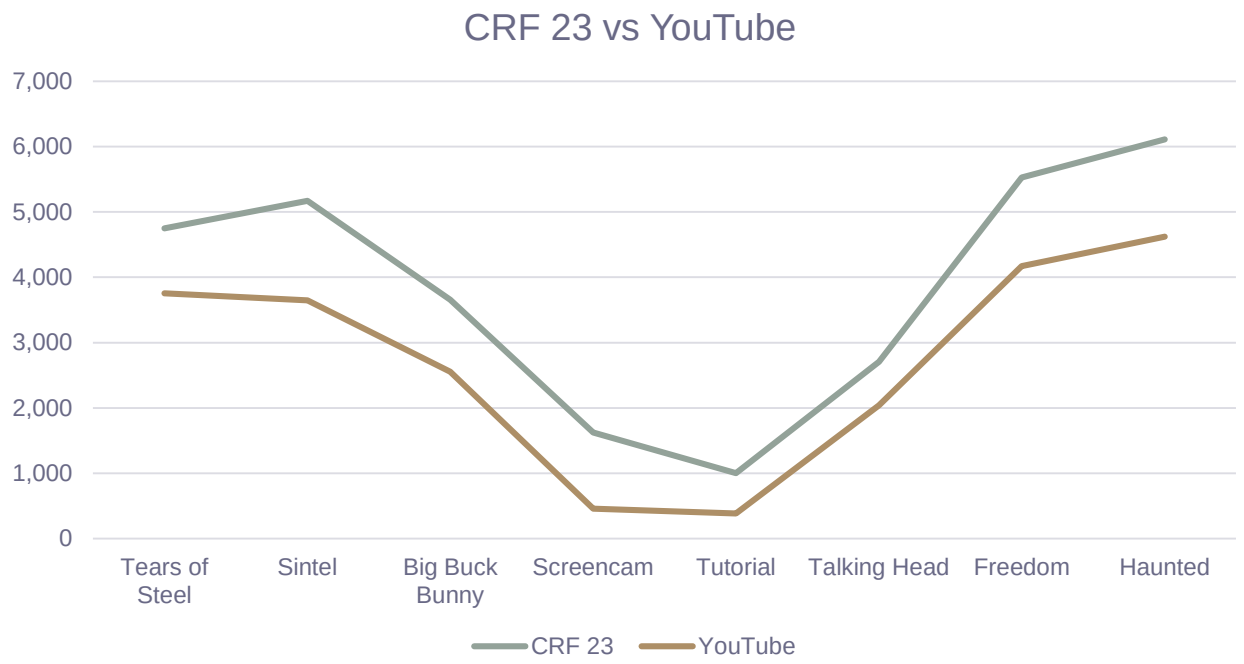
- Our two 24 fps movie-like titles averaged about 4.95 Mbps
- Hollywood titles downloaded from iTunes averaged 5.1 Mbps

- Data rates are similar
- Verifies that CRF 23 and VMAF 93 deliver “Hollywood” quality

VMAF Verification – 93 is the Number

- Real Networks White Paper - VMAF Reproducibility: Validating a Perceptual Practical Video Quality Metric
  - 4K 2D videos
- The results indicate that if a video service operator were to encode video to achieve a ***VMAF score of about 93*** then they would be confident of optimally serving the vast majority of their audience with content that ***is either indistinguishable from original or with noticeable but not annoying distortion.***
  - http://bit.ly/vrqm_5

Reality Check: YouTube Comparison



- Upload files to YouTube; measure data rate
- YouTube uses AI-based per-title optimization
- Pattern very similar

- YouTube averages 1 Mbps lower
- 3 VMAF points lower (1/2 JND)

So

- Full rez 2D videos, CRF 23 = ~93 VMAF = shippable quality
- Significant data point
  - As we'll see – encoding ladder starts at the top

Per-Title Encoding at the Show

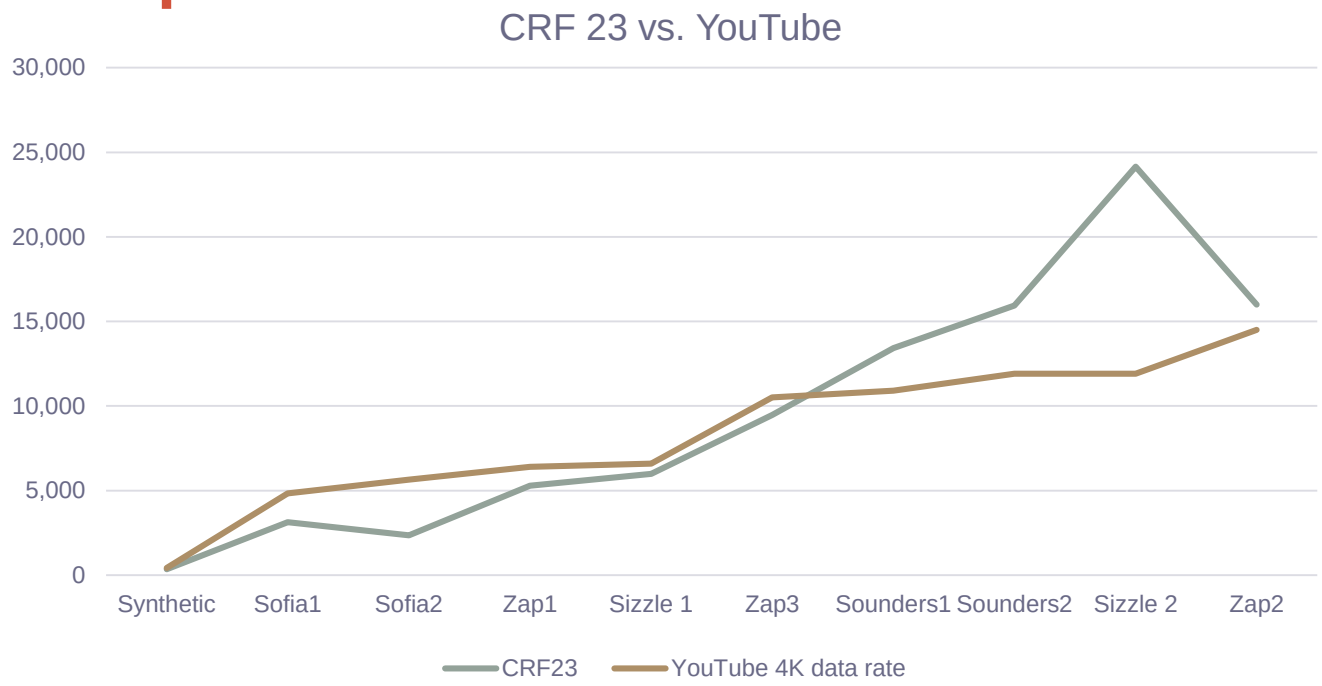
- Capella Systems Cambria Encoder (CRF)
- Bitmovin
- Brightcove
- Elemental (new)
- Harmonic (live!)
- Mux
- Beamr (SDK and cloud)
- ZPEG

What about VR? VR Videos at CRF 23

CRF 23 - H264	Synthetic	Sofia1	Sofia2	Zap1	Sizzle 1	Zap3	Sounders1	Sounders2	Sizzle 2	Zap2
240p	25	56	60	98	86	112	274	311	125	146
360p	41	111	112	191	170	257	552	587	311	358
480p	59	178	173	309	283	452	902	936	625	683
720p	96	353	332	621	590	1,009	1,796	1,902	1,712	1,623
1080p	153	724	639	1,287	1,252	2,321	3,670	4,058	4,968	3,964
2K	214	1,240	1,022	2,175	2,205	3,981	5,981	6,952	10,344	6,863
4K	**355**	**3,129**	**2,348**	**5,286**	**5,987**	**9,467**	**13,431**	**15,934**	**24,159**	**15,999**

- Equirectangular format
  - Ran CRF 23 across multiple resolutions
  - Videos ranged from very simple animations to highly detailed videos
- 4K data rates ranged from 1.15 to 24.1 Mbps
- Per-title absolutely essential to VR

CRF 23 Compared to YouTube



- Similar pattern
- One very major diversion

- CRF 23 averaged about 1.25 Mbps higher
- Remove outlier and delta averaged 25 kbps

Pixvana Verification of VMAF/PSNR

- Create 5 versions of each full rez VR file to be viewed in order
- Center file is CRF 23 value
- Other files vary in intervals of 3 VMAF points
  - File 1 – 87 VMAF
  - File 2 – 90 VMAF
  - File 3 – 93 VMAF
  - File 4 – 96 VMAF
  - File 5 – 99 VMAF
- Tests ~ 20 viewers
  - Choose lowest quality file that’s commercial grade (floor)
  - Choose file at which you see no meaningful improvement (ceiling)

Finding the Floor

Video Name	Average	Standard Deviation	Calculate Data Rate	CRF 23 Data Rate	Delta
Sofia1	1.67	0.71	2,136	3,129	46.49%
Zap1	2.24	1.11	4,056	5,286	30.33%
Sizzle1	2.43	1.05	4,746	5,987	26.15%
Sounders1	2.38	1.33	7,760	13,431	73.08%
Zap3	2.9	0.97	8,750	9,467	8.19%
Average			**5,490**	**7,460**	**35.89%**
Remove outlier			**4,922**	**5,967**	**21.24%**

- CRF 23 averaged 35.89% higher than floor selected by viewers
  - One major outlier
- Was always high, not low
  - Might produce too high a data rate, but in 100% of cases, exceeded floor, so always produced “acceptable” quality

Which Metric? VMAF or PSNR

- VMAF ranged from 90 - 95.5; PSNR from 37.8 - 48.3
- VMAF has much less dispersion and lower standard deviation
- ***Much lower*** Std Dev as percentage of average
- VMAF more accurate than PSNR
- Rule of thumb:
  - CRF 23 s/deliver 93 VMAF or higher
  - If 93 VMAF (again) should be acceptable quality
  - Same for 43.5% PSNR, but less accurate tool

Video Name	Calculate Data Rate	VMAF Calc DR	PSNR Calc DR
Sofia1	2,136	95.5	48.3
Zap1	4,056	93.5	43.6
Sizzle1	4,746	94	45.4
Sounders1	7,760	89.9	37.9
Zap3	8,750	92.0	42.3
Average		**93.0**	**43.5**
Standard Deviation		**2.128**	**3.856**
As percentage of average		**2.29%**	**8.86%**

Finding the Ceiling

Video Name	Average	Standard Deviation	Calc. Data Rate	VMAF Calc DR	PSNR Calc DR
Sofia1	2.52	1.14	2,712	95.90	49.10
Zap1	3.52	0.91	8,880	95.50	46.20
Sizzle1	3.76	1.23	10,560	95.60	47.90
Sounders1	4.1	1.23	13,500	92.60	40.00
Zap3	4.24	0.75	16,480	94.00	44.50
Average			**10,426**	**94.70**	**45.50**

- Go beyond ~95 VMAF or 45.5 PSNR likely not visible by viewers
- Nearly double the data rate for 1.7 VMAF points
  - 1/3 JND

Once You Have Highest it Becomes Math Exercise

- Step 1: Choose highest
- Step 2: Choose lowest
- Step 4: fill in the blanks (between 150/200% apart)

200 kbps

500 kbps

1000 kbps

1600 kbps

2100 kbps

3100 kbps

4600 kbps

Then Question is:

- Netflix approach
  - Compute VMAF scores at multiple resolutions at each data rate
  - Choose best quality at each resolution
  - VMAF proven for 2D by Netflix, what about 3D?

Zap1 - VMAF	4K	2K	1080p	720p	480p	360p	240p
5000	90.19	89.70	84.82				
4500	89.58	88.23	84.38				
4000	88.43	87.50	83.84				
3800	87.88	87.14	83.58				
3600	87.27	86.71	83.25				
3400	86.60	85.72	82.87				
3200	85.80	85.40	82.45				
3000	85.03	85.09	82.01				
2800	83.97	84.34	81.43				
2600	82.86	83.50	80.85				
2400	81.45	82.51	80.09	71.92			
2200	79.79	81.24	79.20	71.35			
2000	77.94	79.82	78.04	70.66			
1800		78.11	76.73	69.70	53.28		
1600		75.91	74.93	68.41	52.82		
1400		73.26	72.64	66.89	52.13	32.07	
1200		69.83	69.69	64.68	51.05	31.75	
1000		65.15	65.75	61.64	49.36	31.17	
900		62.26	63.25	59.64	48.11	30.76	
800		58.69	60.27	57.20	46.54		
700		54.29	56.62	54.06	44.63	29.18	
600		48.79	52.32	50.65	42.02	27.84	
500			46.65	45.96	38.74	25.92	
400			39.06	40.23	34.21	23.11	
300			28.52	32.68	x		
200			13.88	21.73	x		

What about VR

- Ran tests on three files testing top 3 switch points
  - Test different resolutions at that switch point
- Three comparisons
  - Pick best quality or even
  - Round 1 – low res file should win (VMAF 3 higher)
  - Round 2 – should be even (at switch point)
  - Round 3 – high res file should win (VMAF 3 higher)

Clip	Zap1 (dining room/kitchen)		
Encoding complexity	Moderate (CRF 23 = 5,286)		
	VMAF to Subjective	Average	Error
4K to 2K			
Round 1 (2 should win)	Accurate	1.73	
Round 2 (should be tie)	Accurate	1.5	
Round 3 (1 should win)	Accurate	1	
2K to 1080p			
Round 1 (2 should win)	Accurate	1.58	
Round 2 (should be tie)	Accurate	1.45	
Round 3 (1 should win)	Accurate	1.08	
1080p to 720p			
Round 1 (2 should win)	Accurate	1.9	
Round 2 (should be tie)	OK	1.22	Hi Rez
Round 3 (1 should win)	Accurate	1.08	
Low Round	2K-Accurate	2.29	

Overall

- In 2 of 3 trials, worked beautifully (correct 14 out of 15 trials)
- In third trial, incorrect 5 of nine
- But! Highest resolution file always won
  - More testing may be performed, but
  - If close to switch point, go with higher resolution

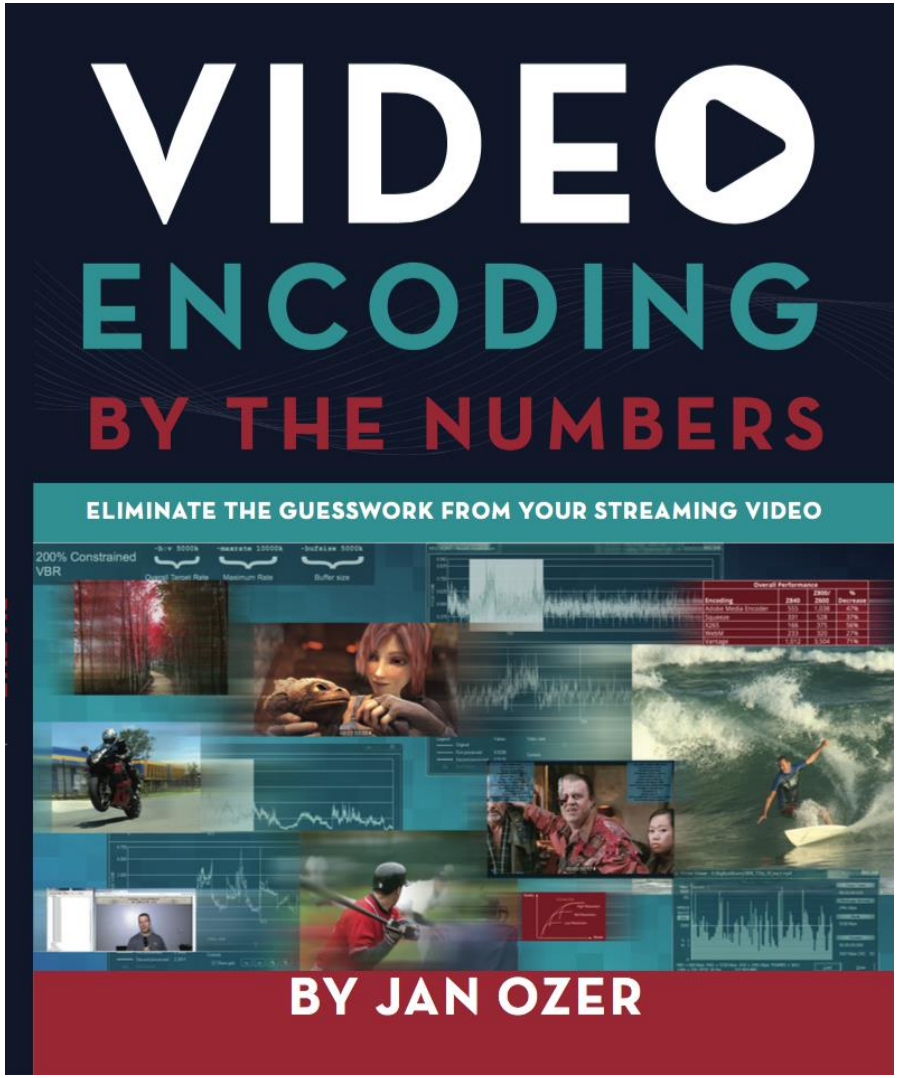
Clip	Sounders 1 (Stadium)		
Encoding complexity	Complex (CRF 23 = 13,431)		
	VMAF to Subjective	Average	Error
4K to 2K			
Round 1 (2 should win)	Inaccurate	1.25	Hi Rez
Round 2 (should be tie)	Accurate	1.42	NA
Round 3 (1 should win)	Accurate	1.17	NA
2K to 1080p			
Round 1 (2 should win)	Inaccurate	1.38	Hi Rez
Round 2 (should be tie)	Inaccurate	1.07	Hi Rez
Round 3 (1 should win)	Accurate	1	NA
1080p to 720p			
Round 1 (2 should win)	Inaccurate	1.17	Hi Rez
Round 2 (should be tie)	Inaccurate	1.15	Hi Rez
Round 3 (1 should win)	Accurate	1.14	NA
Low Round	2K - Inaccurate	1.73	

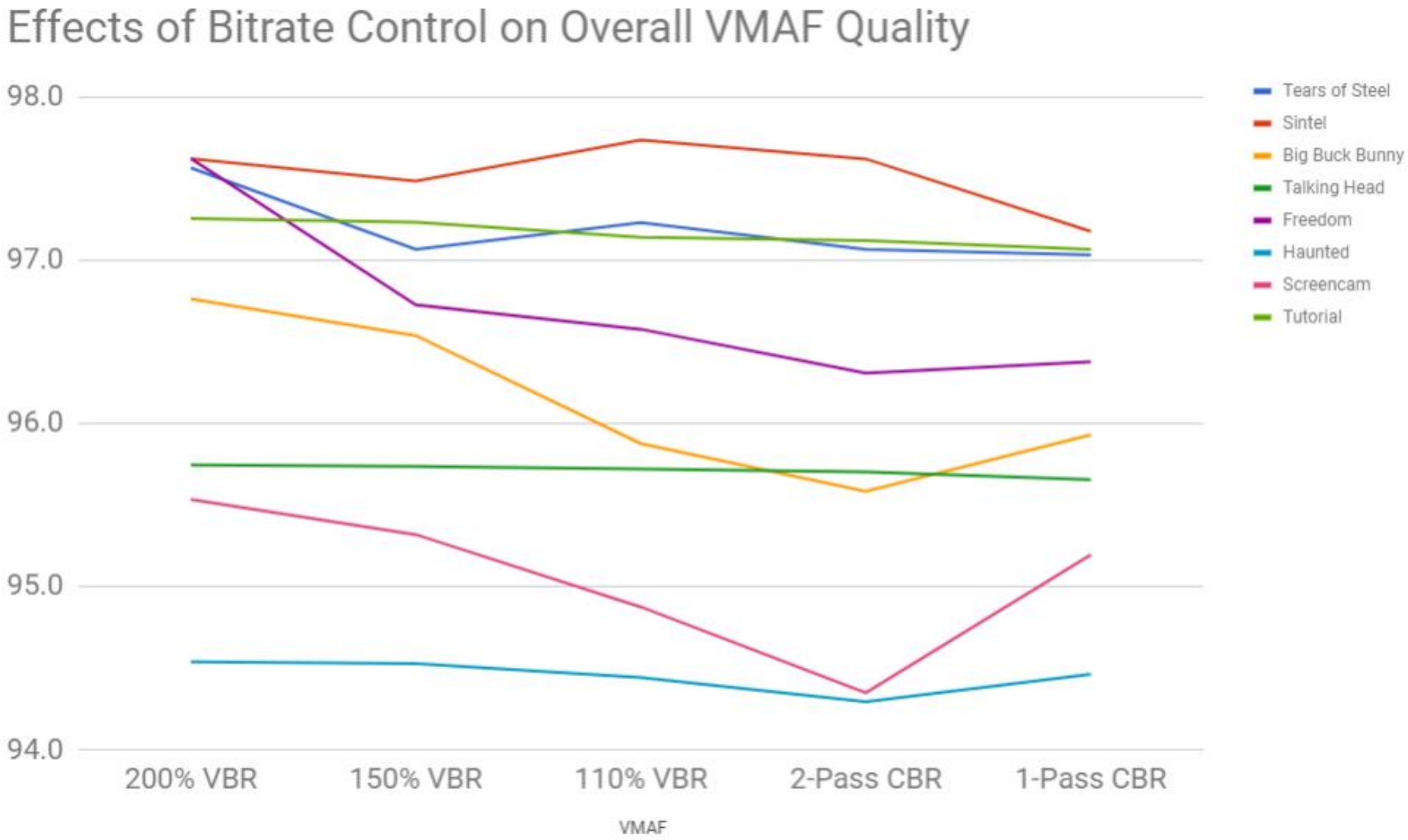
VR – Preliminary Observations

- Different storage formats (equirectangular vs. cube vs. diamond plane) will impact quality at a given data rate more than any encoding parameter or technique
  - Equirectangular appears to lag behind cube mapping (as an example)
- Though VMAF/CRF seem reasonably well proven for equirectangular, haven’t confirmed similar effectiveness for other storage formats

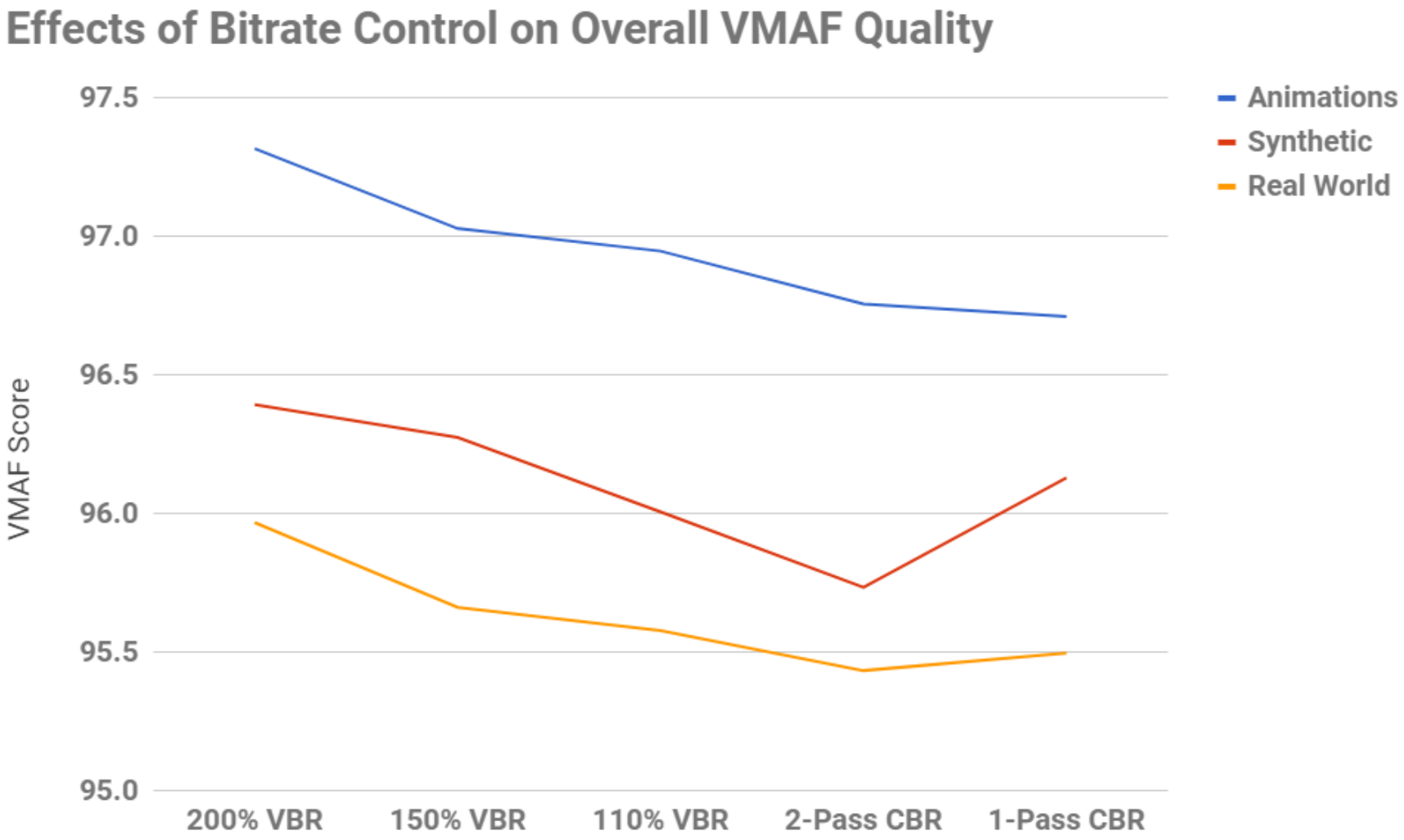
Configuring Your Encodes

- Background: Video Encoding by the Numbers; December 2017
- Eight files
  - 1 movie (Tears of Steal)
  - 2 animations (Sintel, BBB)
  - Two general purpose (concert, advertisement)
  - One talking head
  - Screencam
  - Tutorial (PPT/Video)
- Tied all encoding decisions to PSNR
  - Updating to VMAF/Adding VR now
    - Here – Bitrate control techniques

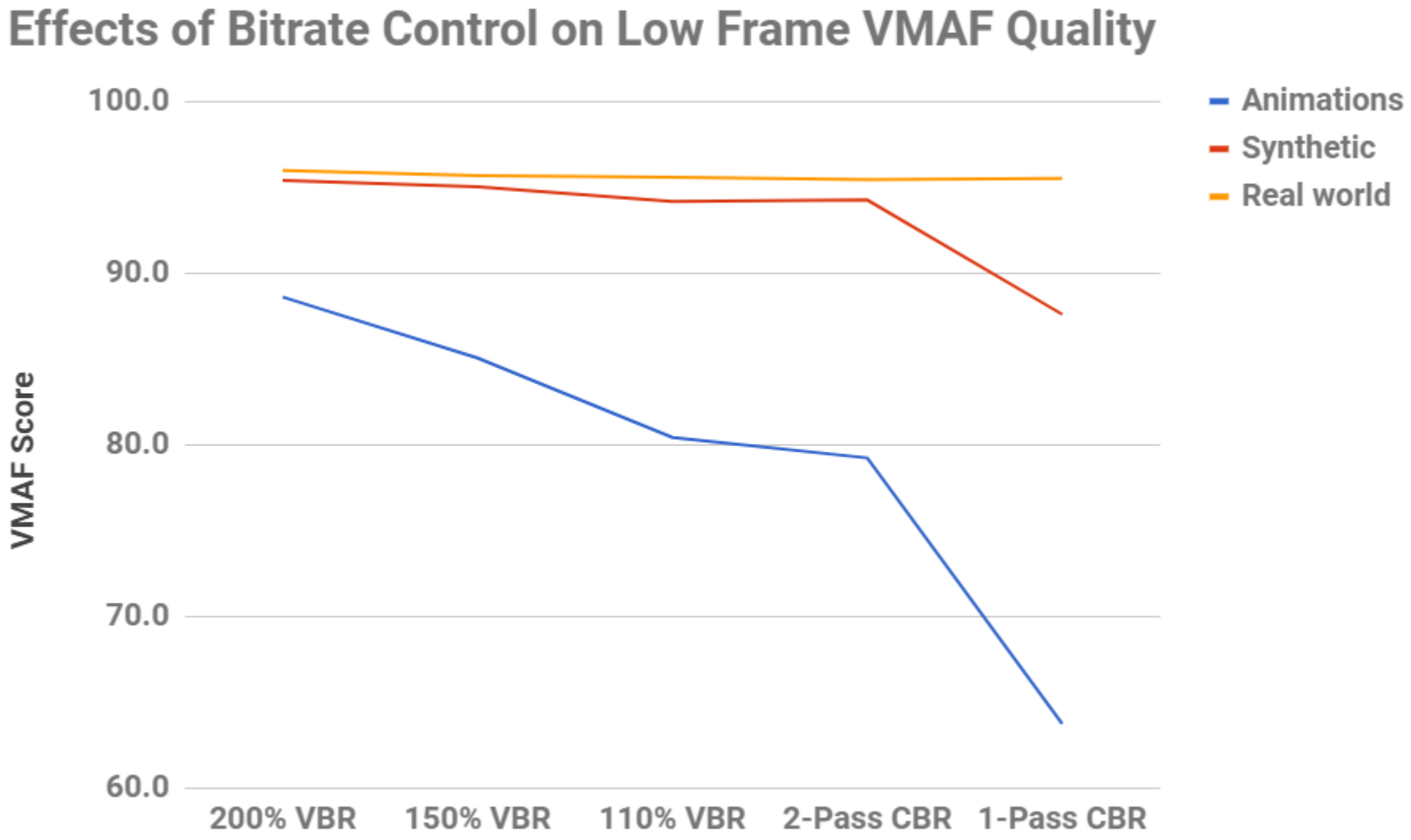




- Not as substantial as you would think



- By class – still not a big deal
  - Over 93 is good enough



- Creates significant issues with low frame quality, particularly in animations

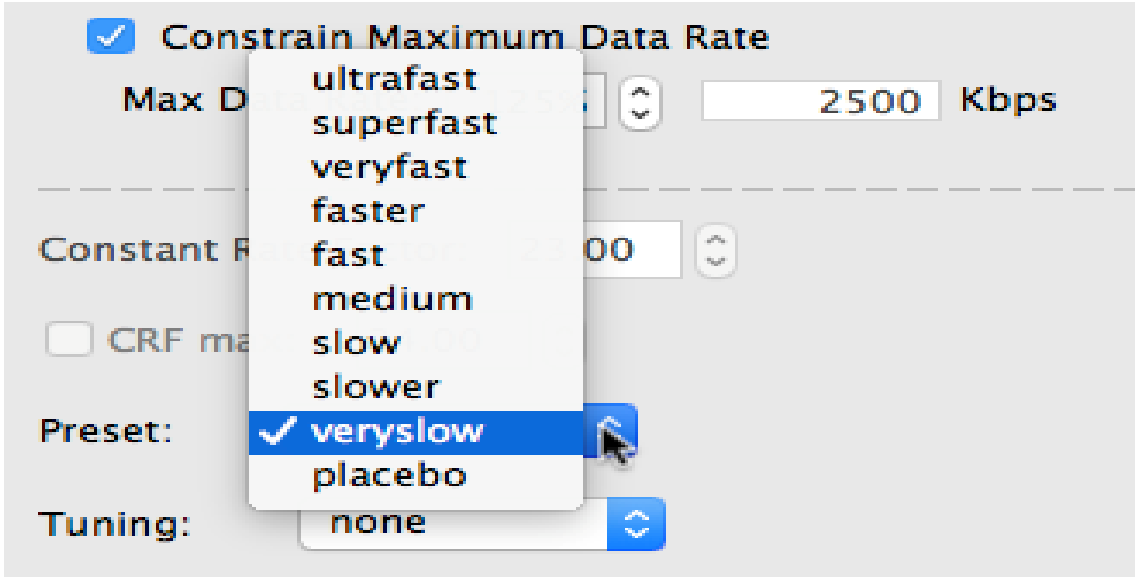
Conclusion

- CRF/VMAF is great combination for choosing data rates and building encoding ladder
- VMAF 93 is the target for acceptable quality for 2D and VR
  - Still early days for VR
- All configuration decisions can and should be measured
- Consider low frame quality as well as overall score

Questions

X264 Preset

- What are presets
  - Simple way to adjust multiple parameters to trade off encoding speed vs. Quality
  - Used by virtually all x264 encoders
  - Medium is generally the default preset



Test Description

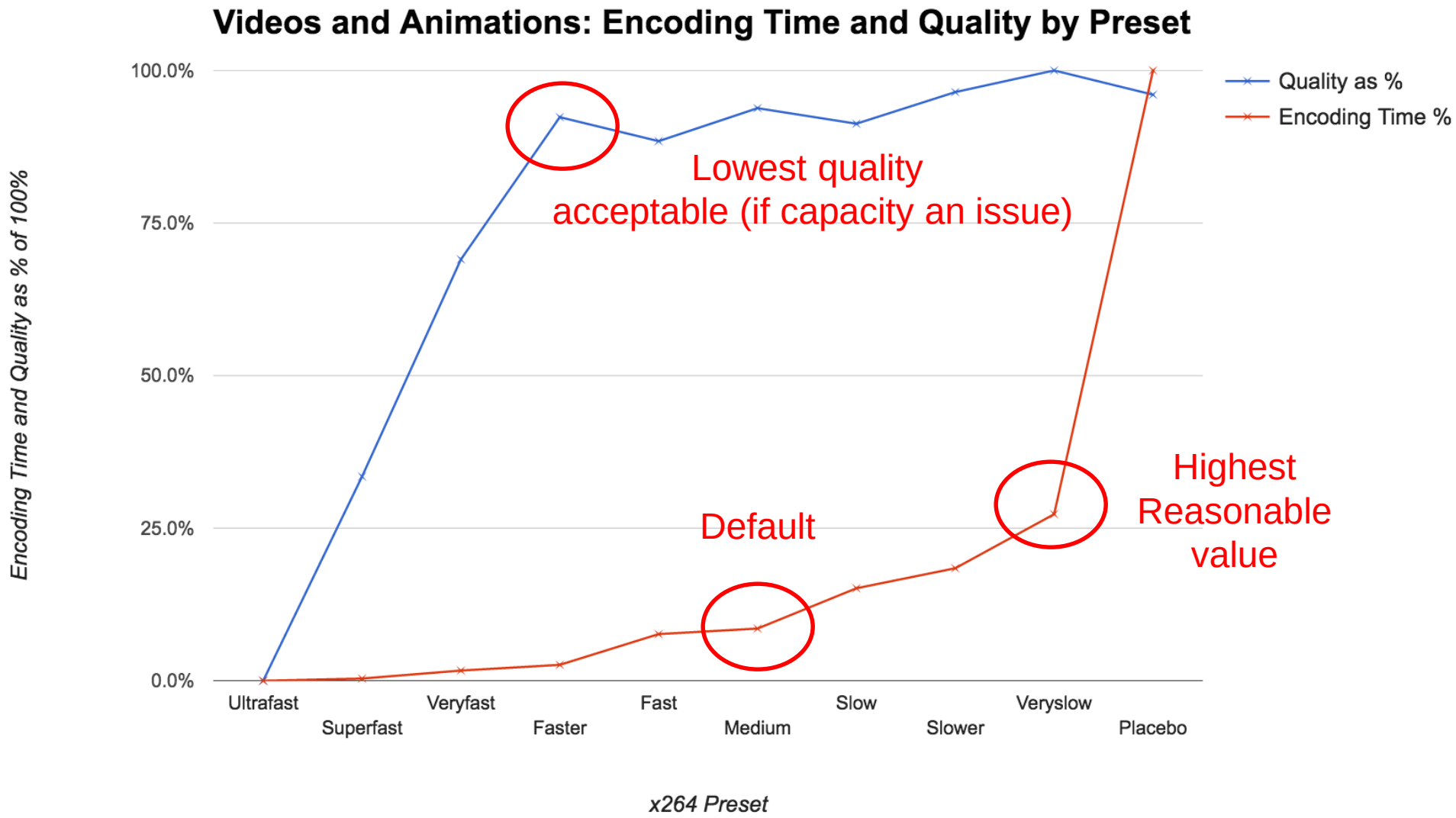
- Eight files
  - 1 movie (Tears of Steal)
  - 2 animations (Sintel, BBB)
  - Two general purpose (concert, advertisement)
  - One talking head
  - Screencam
  - Tutorial (PPT/Video)
- Encode to all presets
- Time encoding
- PSNR

Results Please

	Ultrafast	Superfast	Veryfast	Faster	Fast	Medium	Slow	Slower	Veryslow	Placebo	Total Delta
Tears of Steel	36.07	37.82	38.51	39.23	39.26	39.33	39.27	39.41	39.47	39.40	9.43%
Sintel	35.14	36.71	37.42	38.40	38.43	38.46	38.40	38.55	38.57	38.47	9.75%
Big Buck Bunny	35.19	37.65	38.82	39.49	39.51	39.56	39.50	39.61	39.64	39.54	12.62%
Talking Head	43.38	43.38	44.06	44.39	44.28	44.28	44.21	44.34	44.39	44.29	2.34%
Freedom	38.46	39.26	40.01	40.41	40.32	40.58	40.55	40.69	40.85	40.77	6.22%
Haunted	41.13	41.30	41.89	42.20	42.07	42.27	42.25	42.27	42.35	42.31	2.98%
Screencam	44.46	45.67	46.68	47.12	46.82	46.96	46.95	47.06	46.88	46.76	5.99%
Tutorial	38.47	41.83	43.62	44.50	44.37	44.30	43.99	44.14	44.07	43.91	15.68%
Average	**38.23**	**39.35**	**40.12**	**40.69**	**40.64**	**40.75**	**40.70**	**40.81**	**40.88**	**40.80**	**8.13%**

- Red is lowest quality
- Green highest quality
- Very slow averages best quality
  - But only 8% spread between best and worst

Results Please



Key Frame Interval

	20 sec	10 sec	5 sec	3 sec	2 sec	1 sec	Total Q
TOS	0.936	0.938	0.949	0.964	0.977	1.024	-9.35%
Sintel	0.926	0.932	0.948	0.955	0.969	1.014	-9.59%
Big Buick Bunny	0.525	0.533	0.525	0.541	0.563	0.616	-17.19%
Screencam	0.478	0.478	0.478	0.480	0.493	0.551	-15.09%
Tutorial	0.671	0.673	0.674	0.674	0.675	0.680	-1.25%
Talking Head	0.567	0.569	0.571	0.572	0.569	0.576	-1.72%
Freedom	1.013	1.014	1.014	1.014	1.019	1.022	-0.93%
Haunted	1.665	1.667	1.669	1.669	1.670	1.677	-0.68%

- Encode with interval of 1, 2, 3, 5, 10, 20 second
- Measure quality with VQM
- Green is best, red is worst

- Anyone using keyframe interval of 1 out there?
  - Difference is modest, but why?
- Recommend 3 for ABR (shorter if shorter chunk size)
- Max 10 for other footage

Reference Frames

- What are they?
  - Frames from which the encoded frame can find redundant information
- What's the trade-off?
  - Searching through more frames takes more time, lengthening the encoding cycle
  - Since most redundancies are found in frames proximate to the encoded frame, additional reference frames deliver diminishing returns

How Much Quality?

720p-110CVBR	1 Ref	5 Ref	10 Ref	16 Ref	Max Delta	10 - 16 Delta	16 - 5 Delta
Tears of Steel	39.34	38.99	39.47	39.49	1.28%	-0.04%	-1.26%
Sintel	38.45	38.54	38.58	38.59	0.35%	-0.02%	-0.12%
Big Buck Bunny	38.38	38.48	38.52	38.51	0.36%	0.03%	-0.08%
Talking Head	44.27	44.36	44.39	44.40	0.29%	-0.03%	-0.10%
Freedom	40.68	40.80	40.85	40.87	0.47%	-0.06%	-0.19%
Haunted	42.24	42.32	42.35	42.36	0.26%	-0.02%	-0.08%
Average - 720p	**40.56**	**40.58**	**40.69**	**40.70**	**0.34%**	**-0.02%**	**-0.30%**

- 16 is best
  - Miniscule difference between 16 and 10 (.02%)
  - .3% delta between 5 and 16

How Much Time?

Encoding Time	1 Ref	5 Ref	10 Ref	16 Ref	Max Delta	10 - 16 Delta	16 - 5 Delta
Tears of Steel	39	49	72	91	133%	-21%	-46%
Sintel	40	53	71	76	90%	-7%	-30%
Big Buck Bunny	41	53	68	85	107%	-20%	-38%
Talking Head	37	47	61	77	108%	-21%	-39%
Freedom	99	142	200	263	166%	-24%	-46%
Haunted	47	65	93	123	162%	-24%	-47%
Average - 720p	**51**	**68**	**94**	**119**	**136%**	**-21%**	**-43%**

- 16 is ~ 2.5 x longer than 1 reference frame
  - Cutting to 5 reduces encoding time by 43% (close to doubling capacity)
  - Reduces quality by .3%

Reference Frames

- Recommend 5 as best blend of performance and quality
  - Can increase encoding capacity by ~40% over 16 with no discernable impact on quality

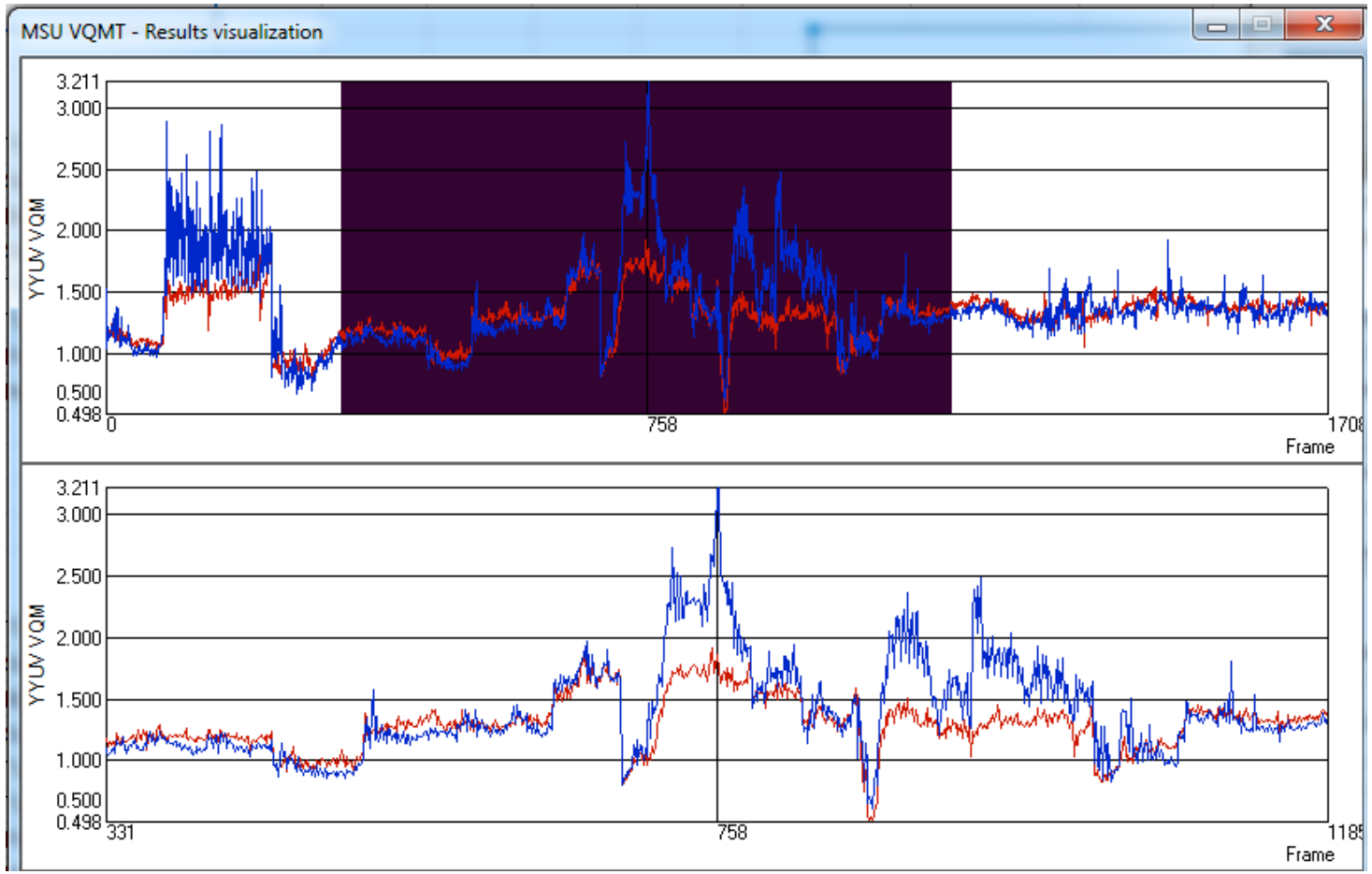
VBR or CBR?

	200% VBR	150% VBR	125% VBR	CBR 2Pass	CBR 1Pass	Total Quality Delta
TOS	1.278	1.278	1.297	1.379	1.507	-18%
Sintel	1.211	1.212	1.209	1.306	1.439	-19%
Big Buick Bunny	0.994	0.995	0.996	1.073	1.164	-17%
Screencam	0.480	0.485	0.501	0.654	0.696	-45%
Tutorial	0.845	0.845	0.845	0.869	0.850	-1%
Talking Head	0.561	0.562	0.561	0.582	0.621	-11%
Freedom	1.620	1.618	1.621	1.639	1.682	-4%
Haunted	1.669	1.665	1.667	1.676	1.710	-2%

- Encode using 200%, 150, and 125% constrained VBR; 1 & 2 pass CBR
- Measure quality with VQM
- Green is best, red worst

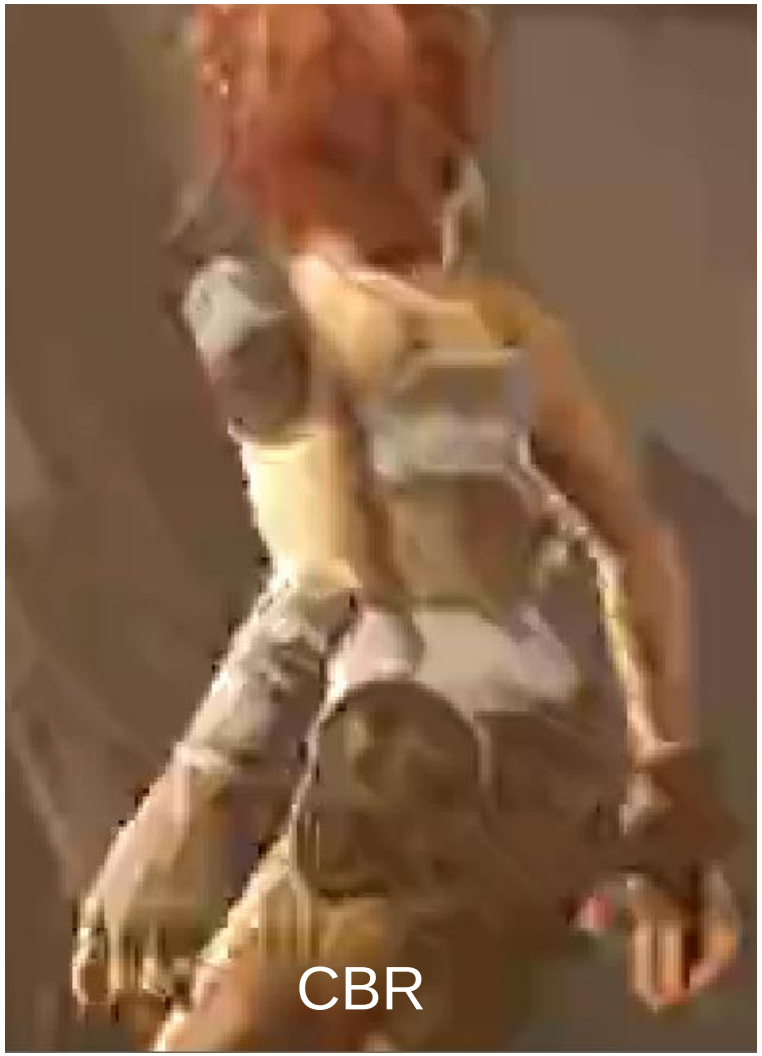
- It gets even worse

Some Files will Show Quality Glitches



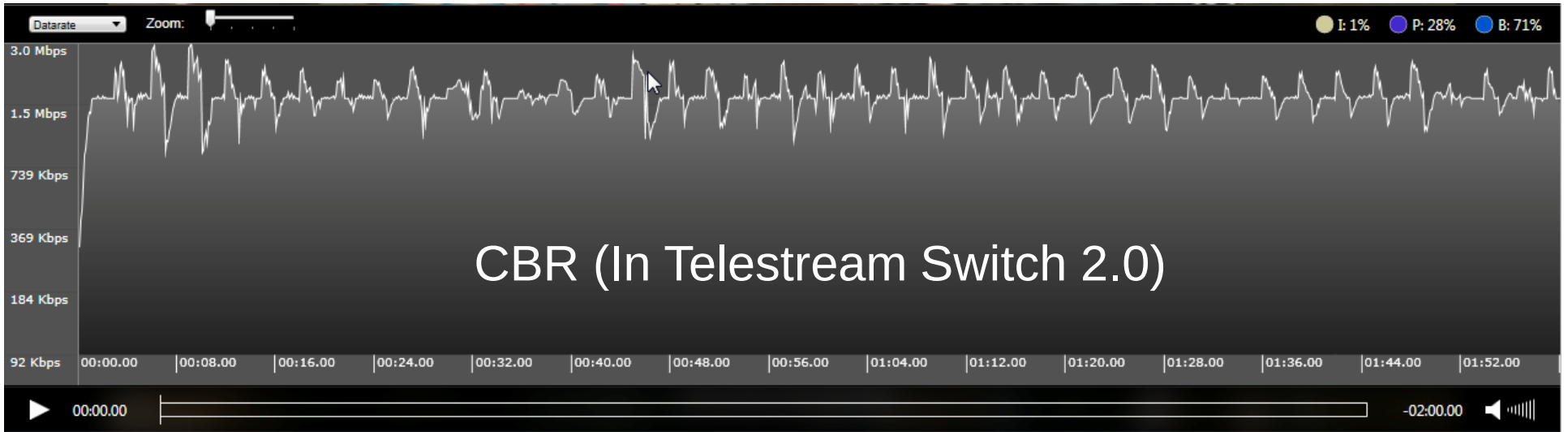
Files very close most of the time with notable exceptions

Transient Quality Issues



Definitely Can Be Smoothness Issues

Supposed to be 125% constrained



CBR vs VBR

- Big issue:
  - Overall quality
  - Transient quality
- Deliverability is a huge issue with VBR
  - http://bit.ly/VBR_CBR_QOE
- I recommend 110% constrained VBR; best blend of *quality* and *deliverability*

Building Your Encoding Ladder

- Step 1: Choose lowest rate for mobile
- Step 2: Choose highest supported data rate (cost issue)
- Step 3: Choose data rate around 3 mbps (highest sustainable)
- Step 4: fill in the blanks (between 150/200% apart)

200 kbps

500 kbps

1000 kbps

1600 kbps

2100 kbps

3100 kbps

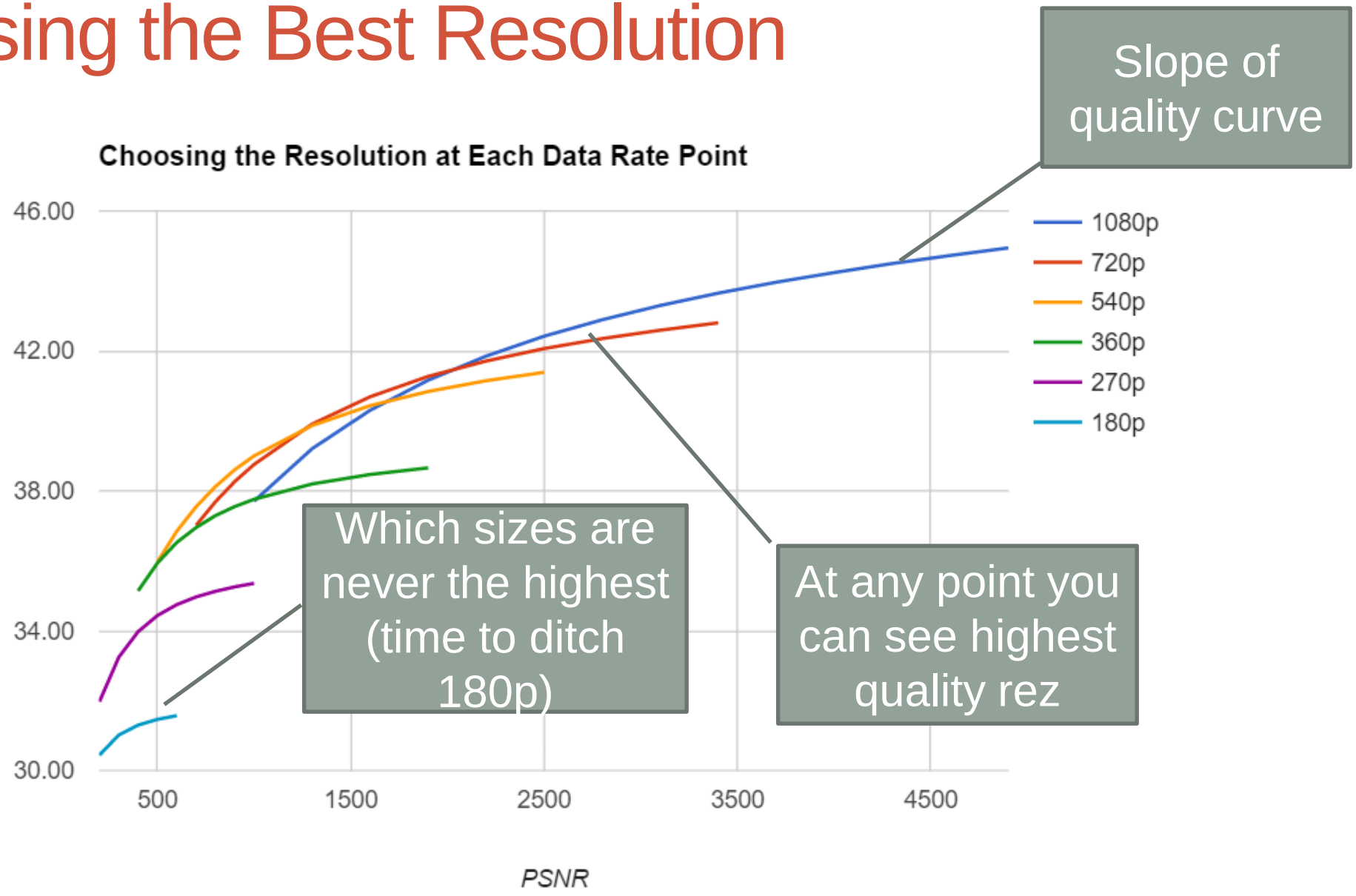
4600 kbps

Then Question is:

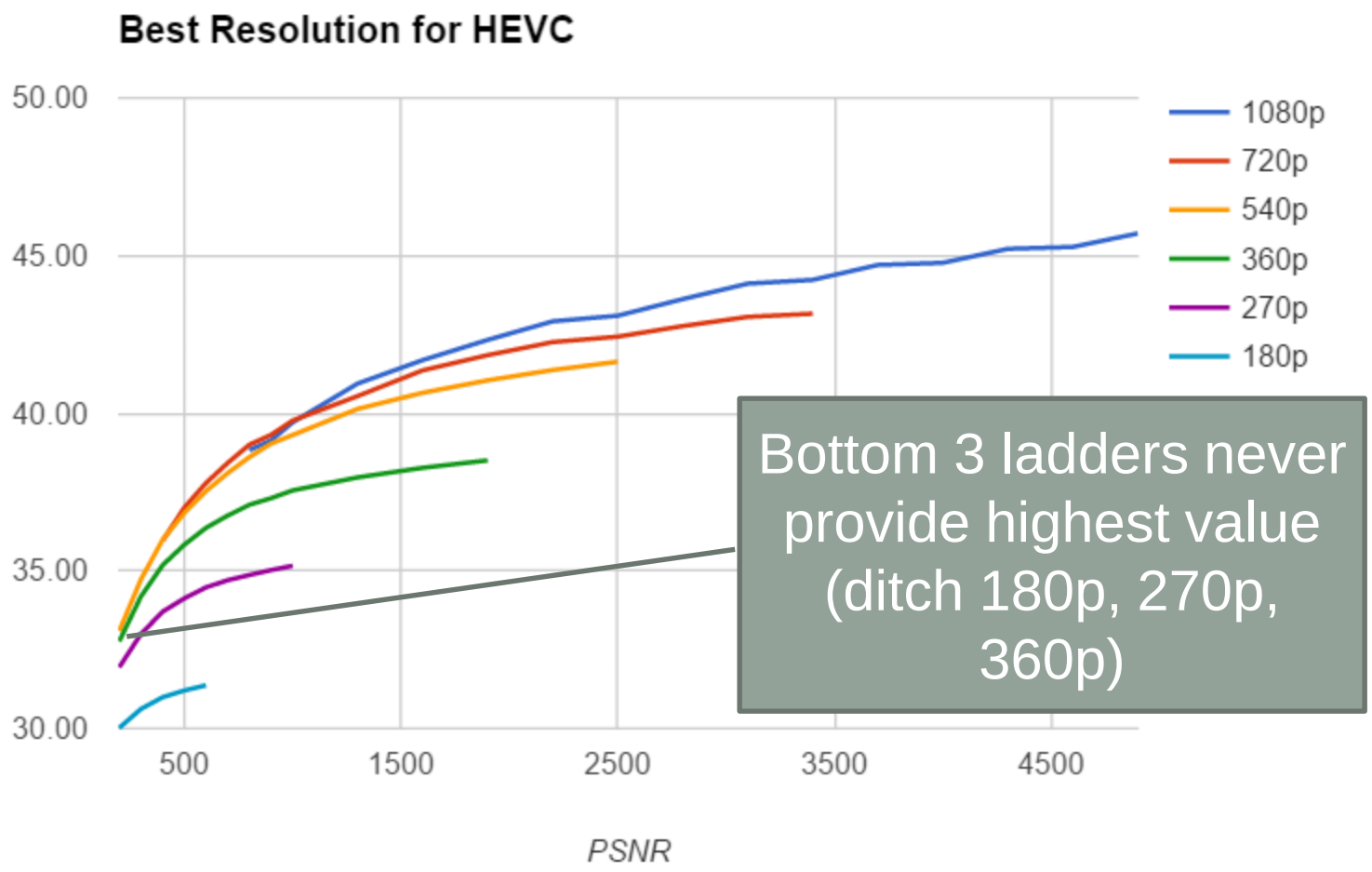
- Best resolution at each data rate
- Similar to per-title approach used by Netflix

PSNR	1080p	720p	540p	360p	270p	180p
4900	44.94					
4600	44.73					
4300	44.50					
4000	44.24					
3700	43.96					
3400	43.65	42.80				
3100	43.30	42.59				
2800	42.89	42.35				
2500	42.42	42.07	41.39			
2200	41.85	41.71	41.15			
1900	41.16	41.27	40.84	38.65		
1600	40.30	40.69	40.43	38.47		
1300	39.20	39.91	39.87	38.20		
1000	37.70	38.75	39.00	37.76	35.35	
900		38.27	38.60	37.55	35.25	
800		37.69	38.12	37.29	35.13	
700		37.01	37.54	36.95	34.97	
600			36.85	36.52	34.74	31.57
500			35.97	35.93	34.43	31.47
400				35.14	33.97	31.30
300					33.24	31.02
200					31.97	30.44

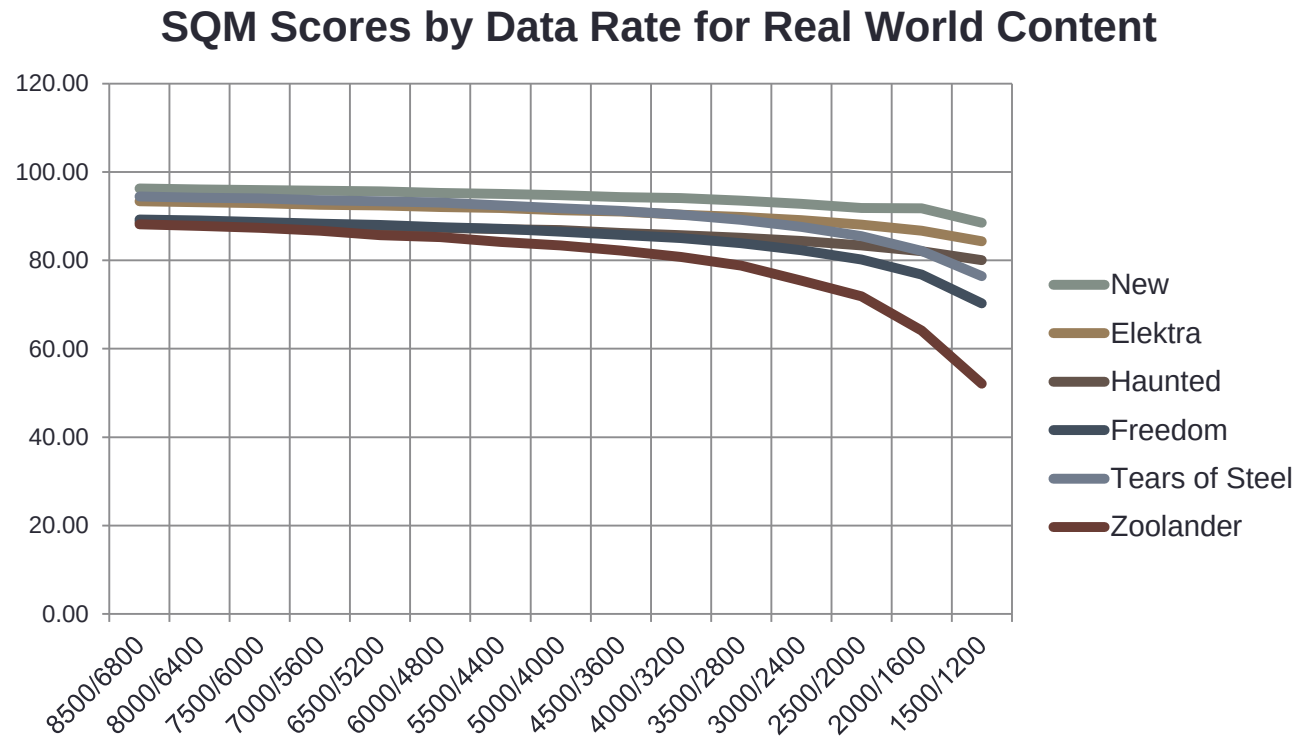
Choosing the Best Resolution



Choosing the Best Resolution HEVC



How Low Can You Go?



- SQM – Higher is better
- Here we see Zoolander drop below 80 right around 4 mbps
- Others stay in excellent range throughout

What About Animation?

	SQM	
VQM	Real World	Animated
8500/6800	91.71	92.84
8000/6400	91.48	92.59
7500/6000	91.19	92.38
7000/5600	90.84	92.19
6500/5200	90.49	92.06
6000/4800	90.10	91.68
5500/4400	89.63	91.40
5000/4000	89.12	90.88
4500/3600	88.49	90.33
4000/3200	87.72	88.77
3500/2800	86.74	88.83
3000/2400	85.28	87.62
2500/2000	83.50	85.92
2000/1600	80.62	83.26

SQM Level	Real World Data Rate	Animated Data Rate	Delta
91.71/91.68	8500	6000	2500
90.84/90.88	7000	5000	2000
90.10/90.33	6000	4500	1500
87.72/87.62	4000	3000	1000

- Animated scores achieved similar quality levels to real world at much lower data rates
- Should be able to produce the same quality on animated content at a much lower data rate

To Run These Tests

Overall Performance			
Analysis	Z840	Z800	% Decrease
Convert to YUV	56	367	85%
MSU VQMT	860	1,701	49%

- Computer/disk speed matters
- Use the fastest computer you have
- Use an SSD drive if at all possible
- HP Z840 have been awesome for me

Questions?

- Questions