



APRIL 5-9, 2025 (EXHIBITS APRIL 6-9)
LAS VEGAS, NV

Improving Streaming QOE Analytics

By Using Timed Metadata and
Encoder-Reported Quality Scores

Yuriy Reznik
Brightcove, Inc.

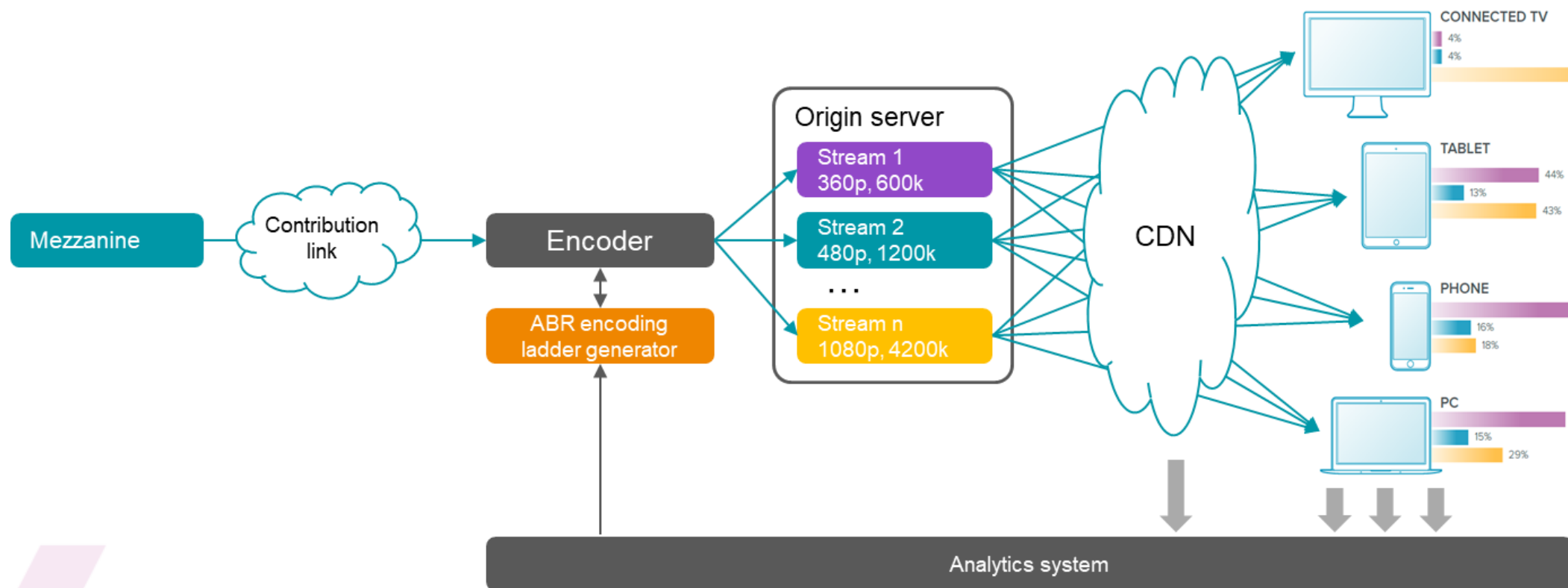
Outline

- Streaming analytics 101
 - Related standards: CTA 2066, ITU-T P.1203
 - Challenges
- Adding full-reference quality metrics
 - Carriage of metrics
 - Accounting for characteristics of user devices & viewing setups
 - Fusion with playback statistics
- Example of implementation
- Conclusions

Streaming analytics 101

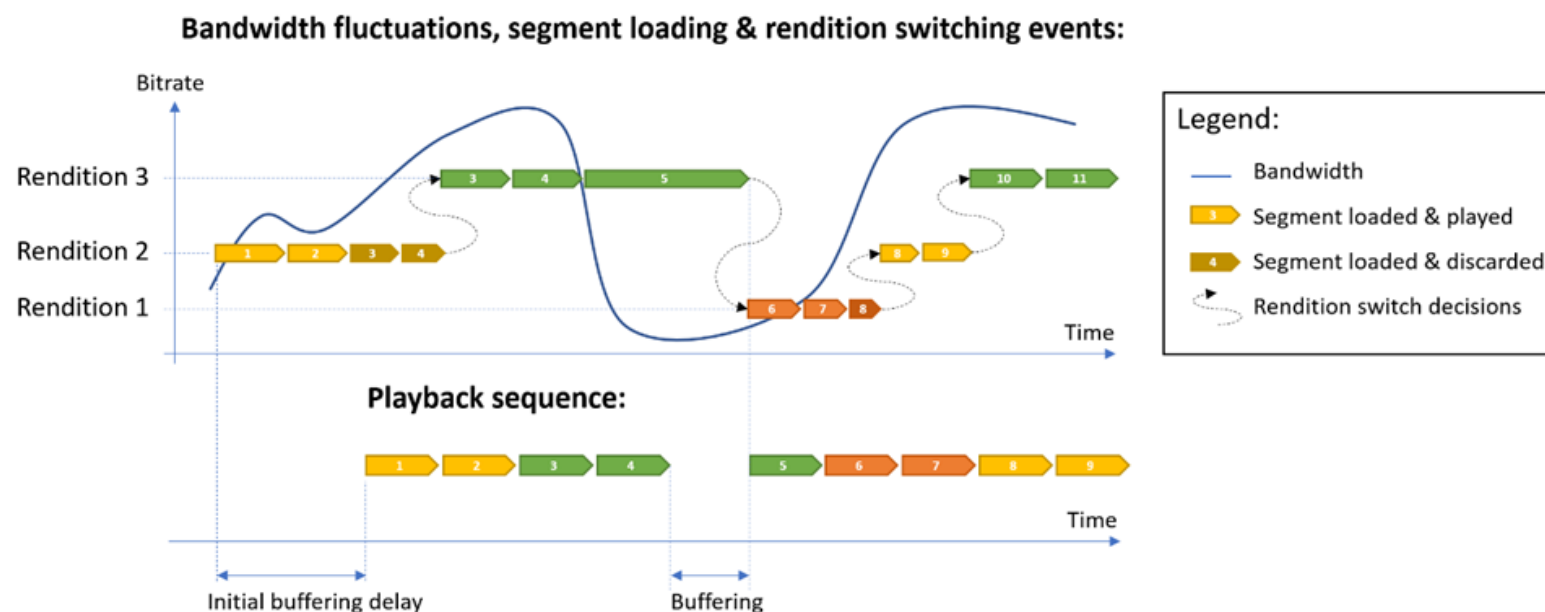
ABR Streaming System

Overall architecture:



Streaming Session

Example:

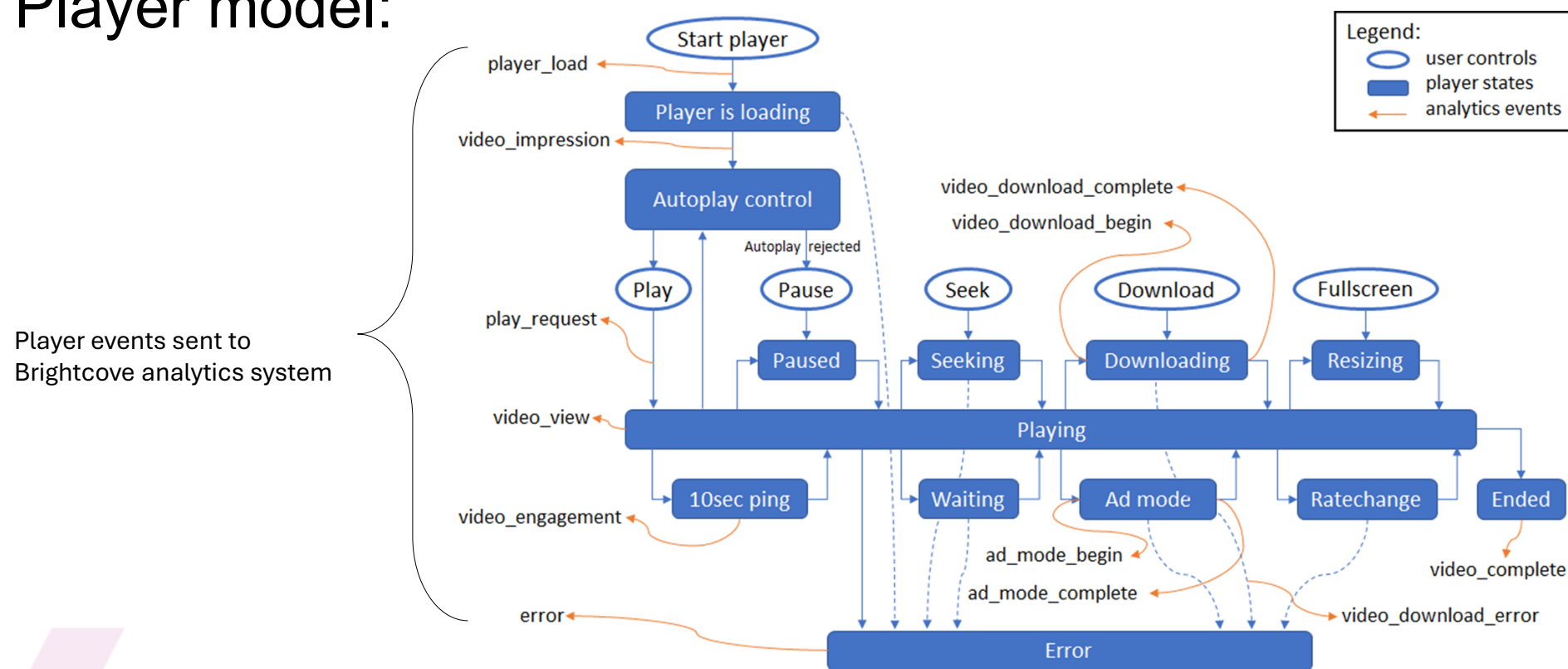


Known effects:

- Playback starts with delay
- Players can buffer mid-session
- Try to minimize buffering by selecting bitrates adaptively

Streaming Analytics Data Collection

Player model:

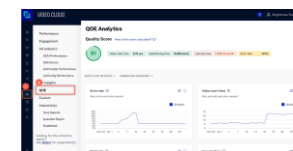


Player-reported metrics & events

Category	Parameter	Description
session	session	A unique number assigned to each session
	seq	Sequential number of an event within a session
client	device_type	Device type. Possible values: "desktop", "mobile", "tablet", "tv", "other"
	device_os	OS type. Possible values: "windows", "osx", "linux", "android", "ios", "webos", "other"
	browser	Browser type. Possible values: "chrome", "firefox", "safari", "edge", "ie", "opera", "other"
	player	Player type. Possible values: "app" – dedicated application, "web" – JS / browser-based player
	player_width	Player window width [pixels]
	player_height	Player window height [pixels]
rendition	rendition_indicated_bps	Rendition bitrate [bps]. Sum of audio and video bitrates.
	rendition_width	Video width as encoded [pixels]
	rendition_height	Video height as encoded [pixels]
	rendition_framerate	Video framerate [fps]
	video_codec	Video codec type. Possible values: "h264", "hevc", "av1"
	video_codec_profile	Video codec profile. Possible values: "baseline", "main", "high"
	format	Streaming format. Possible values: "hls_v3", "hls_v7", "dash"
	segment_duration	Segment duration [seconds]
playback	video_seconds_viewed	Seconds of media content played in the period between the last two player events
	forward_buffer_seconds	The number of seconds of media content buffered but not yet played
	rebuffering_seconds	The total number of seconds the player was "buffering" in the period between the last two player events
	rebuffering_count	The number of times the player was "buffering" in the period between the last two player events
	media_bytes_transferred	The total number of bytes transferred since the start of the session
network	measured_bps	Network bandwidth [bps] estimated based on size and delivery time of the last segment downloaded

<https://apis.support.brightcove.com/data-collection/getting-started/overview-data-collection-api-v2.html>

Final analytics reports



Summary by Device

Title	Views ↓ ⓘ	Error rate ⓘ	Video start time ⓘ	Video resolution ⓘ	Rebuffering time ⓘ	Stall rate ⓘ	Bitrate ⓘ	Upscale time ⓘ
All	16,491	2.4%	1.769		19:42:54	26.38	1.20 Mbps	00:01:18
desktop - Windows 10	8,655	2.4%	1.575		34:27:30	26.51	1.35 Mbps	00:01:30
desktop - OS X	4,138	1.6%	1.318		11:18	61.83	779.22 Kbps	00:09:07
desktop - Linux	1,900	3.2%	0.39		31:42	234.03	718.68 Kbps	00:07:25
mobile - Android	1,113	6.7%	5.089		51:19	122.08	1.21 Mbps	00:02:28
mobile - iOS	395	2.7%	3.338		09:50	76.52	60.05 Kbps	00:00:25

<https://studio.support.brightcove.com/analytics/module/qoe-insights.html>

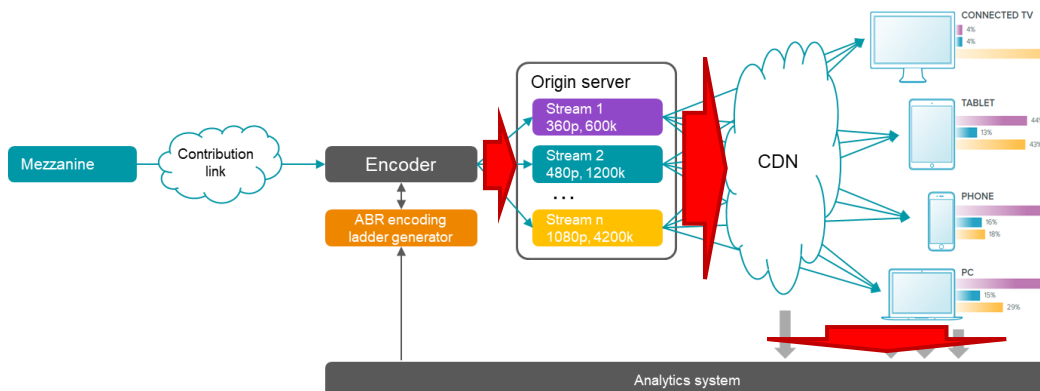
Standards & limitations

- Relevant standards
 - CTA 2066 – Streaming QOE events, properties, and metrics, 2019
 - ITU-T P.1203 -- Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport, 2017
- Limitations:
 - By collecting only stall/bitrate/resolution metrics, we can't make accurate predictions about the quality of video encoding and hence quality of experience
 - P.1203 infers the amount of codec-introduced noise by bitstream parsing, but it only brings an approximate assessment
 - A more precise assessment of QOE can benefit from **knowledge of Full-Reference quality metrics**, such as PSNR, SSIM, PQR, VMAF, etc.

Adding Full-Reference Metrics

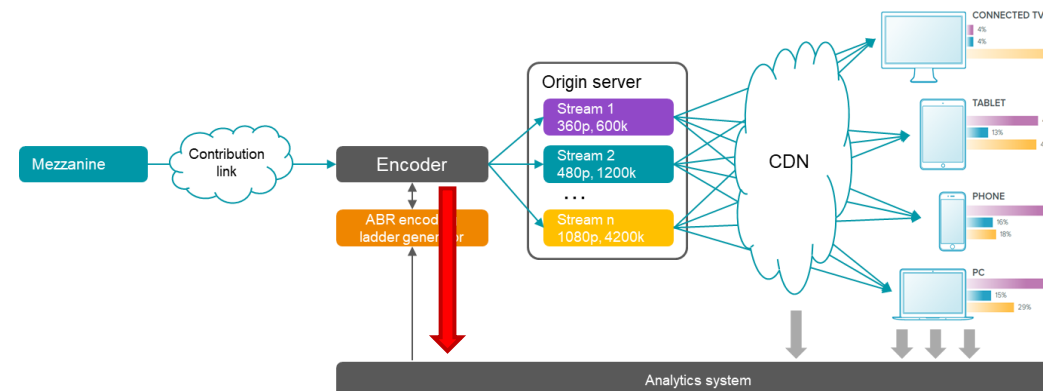
Carriage of encoder quality metrics

Downstream signaling:



- Pros: complementary to existing systems
- Cons: requires updates of most components in the streaming chain

Encoder analytics:



- Pros: minimal work needed
- Cons: the analytics system must be able to associate encoder and playback data (not a problem in an OVP)

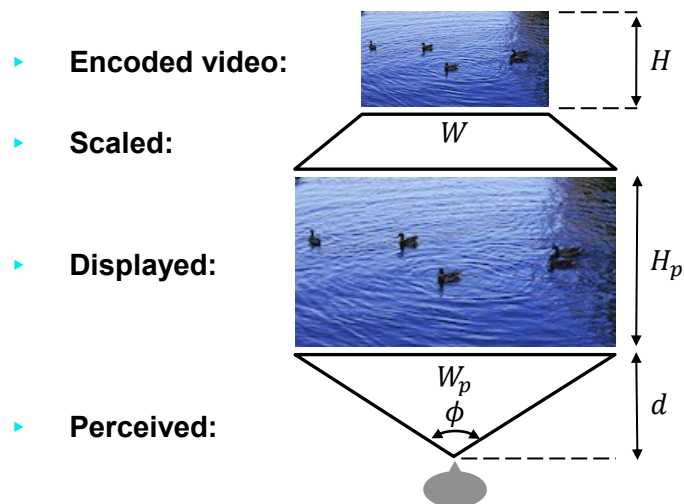
Downstream signaling

- Using timed metadata
 - MPEG-C Part 10 aka ISO/IEC 23001-10:2020
 - Allows embedding of quality metrics in ISOBMFF streams
- Manifest-level signaling
 - Possible using HLS SCORE attributes
 - Cons:
 - Only sequence level granularity is supported in HLS
 - No consistent mechanism across HLS and DASH
- CMSD MQA signaling
 - New mechanism under development in CTA WAVE and SVTA
 - U. Pal and W. Law, Proposal for CMSD-based transmission of Media Quality Assessment (MQA) data, 27 Nov 2024
 - Example: CMSD-Static: mqa= ("VMAF";v=96 "PSNR";v=38)

Mapping Full-Reference Metrics to MOS scores on different devices

Known science

Video reproduction chain



Main parameters involved

Parameters	Meaning	Unit
W, H	encoded video width, height	pixels
W_p, H_p	display/player width, height	pixels
d	viewing distance	inches
ρ	display pixel density	dots per inch
$\phi = 2 \arctan\left(\frac{W_p}{2d\rho}\right)$	viewing angle	degrees
$\phi_c = 2 \arctan\left(\frac{W_p/W}{d\rho}\right)$	angle to 2 pixels (1 cycle)	degrees
$u = \frac{1}{\phi_c}$	angular resolution of video	cycles per degree (cpd)

Relevant for human perception

- viewing angle ϕ → angular span of video frame, as visible on screen
- angular resolution u → inverse of angular span of 2 pixels (length of smallest "cycle") in encoded video

Reproduction on different devices

Viewing setup parameters

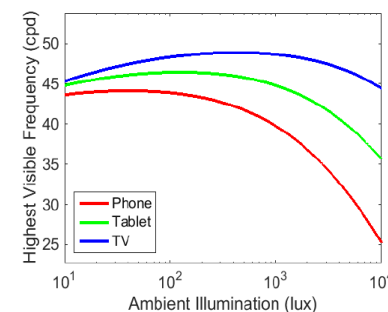
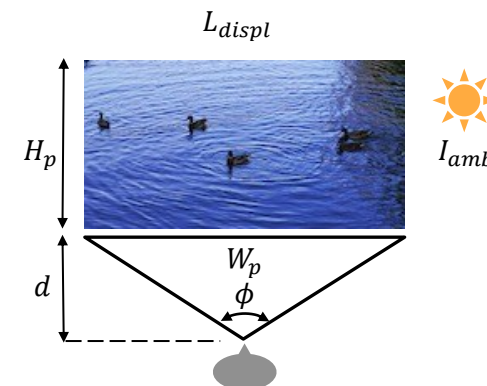
Device	Viewing distance	Display size	Brightness	Ambient	Background
TV	2-6H, med=3H	32-80", med=50"	400 nits	50-200 lux	10-30% reflective
PC / laptop	12-30", med=24"	13-36", med=20"	200 nits	100-500 lux	Varies
Tablet	10-22", med=18"	7-12", med=9"	200-400 nits	50-500 lux	Varies
Phone	7.5-22", med=14"	4-6", med=5.5"	100-300 nits	10-10000 lux	Varies

Angular metrics

Device	Viewing angle	Angular resolutions when rendering video full screen					Max. visible resolution*
		360p	540p	720p	1080p	4k	
TV	33.0 degrees	9.4 cpd	14.3 cpd	18.9 cpd	28.3 cpd	56.5 cpd	48.3 cpd
PC/Laptop	40+ degrees	7.7 cpd	11.5 cpd	15.4 cpd	23.1 cpd	46.1 cpd	46.2 cpd
Tablet	24.6 degrees	12.8 cpd	19.2 cpd	25.6 cpd	38.4 cpd	76.9 cpd	46.2 cpd
Phone	18.2 degrees	17.9 cpd	26.9 cpd	35.9 cpd	53.8 cpd	107.6 cpd	44.1 cpd

Observations

- ▶ Viewing angles and angular resolutions are very different on different devices!
- ▶ With high-resolution content and small form-factor devices, angular resolutions may exceed maximum resolutions visible by human eye!



(*) L. Kerofsky, R. Vanam and Y. Reznik, "Adapting Objective Video Quality Metrics to Ambient Lighting," QOMEX 2015.

Perceived quality

Westerink & Roufs experiments (1989)*

- ▶ Controlled environment, 20 subjects, 5 images, 0-10 categorical scale
- ▶ Varied: viewing distance, resolution, and picture size

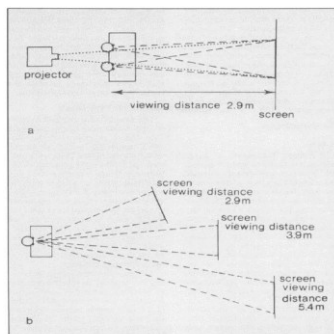


Figure 1. Experimental configurations: (a) Experiment 1; (b) Experiment 2.

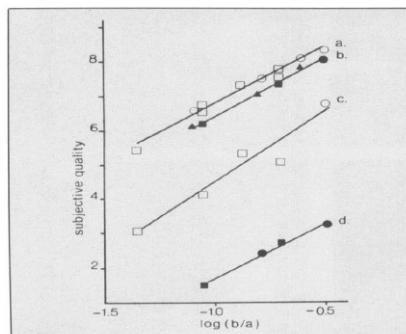


Figure 4. Subjective image quality as a function of the picture angle. The subjective quality values on the vertical axis are plotted as a function of $\log(b/a)$, which differs from the picture angle ϕ only by a constant. The different sets correspond to different resolutions: (a) greater than 33° , (b) between 23 and 33° , (c) between 8.6 and 23° , and (d) between 2.6 and 8.7° . Different symbols represent different viewing distances: $\square = 2.9$ m, $\triangle = 3.9$ m, $\diamond = 5.4$ m. Every point is the result of 80 judgments.

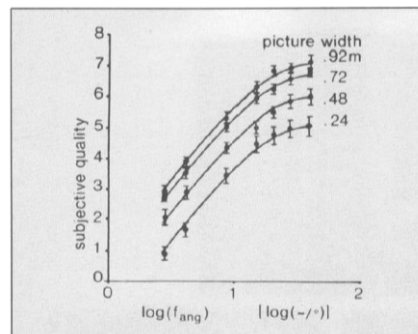


Figure 3. Subjective quality as a function of resolution. Every point is the result of 100 judgments, and the error indicated is plus or minus the standard error of the mean.

(*) J. Westerink and J. Roufs, "Subjective image quality as a function of viewing distance resolution and picture size," SMPTE Journal, vol. 98, 1989, pp. 113-19.

(**) P. G. J. Barten, "Effect of picture size and definition on perceived image quality," IEEE Trans. Electron. Devices, vol. 36, no. 9, pp. 1865-1869, Sept. 1989.

Observed phenomena:

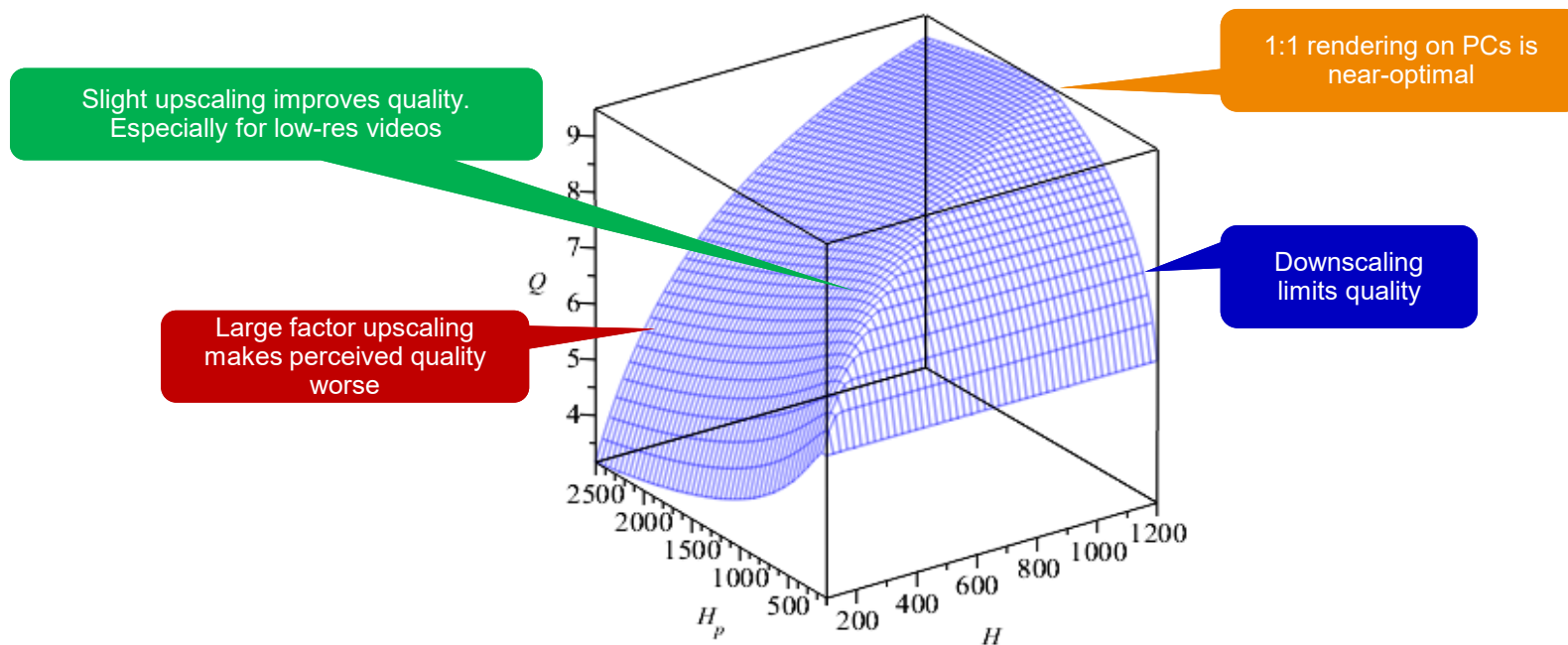
- ▶ Perceived quality grows approximately as logarithm of viewing angle (ϕ)
- ▶ Perceived quality also grows with angular resolution (u), but saturates at around 25-40 cycles/degree

Classic model describing these effects*

$$Q_{WR}(\phi, u) = 3.6 \log(\phi) + 2.9 + 4.6 \log(u) + 2.7 \log(u)^2 - 1.7 \log(u)^3$$

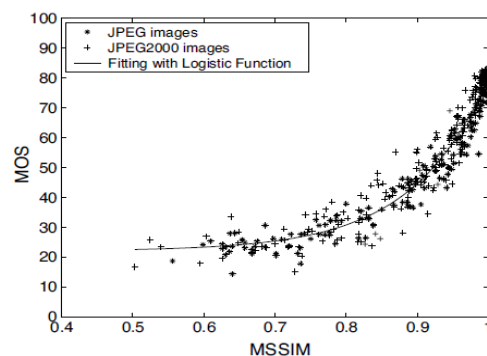
Example: effects of scaling / PC viewing

WR model for PC viewing ($d=24, p=96$)

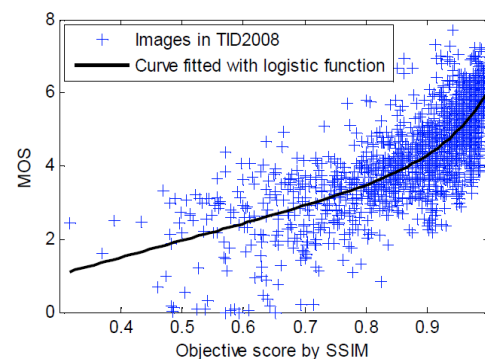


Codec noise and perceived quality

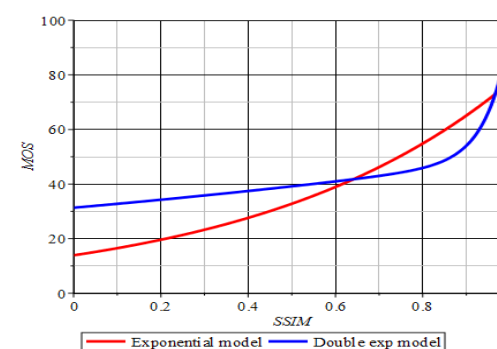
For SSIM* metric this relationship is well studied



(*) Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, "Image quality assessment: from error visibility to structural similarity". IEEE Transactions on Image Processing 13 (4) (2004).



L. Zhang, L. Zhang, X. Mou, D. Zhang, "FSIM: a feature similarity index for image quality assessment", IEEE Trans Image Process. 20 (8) (2011)



U. Engelke, M. Kusuma, H-J. Zepernick, M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging". Signal Processing - Image Communication, 24 (7) (2009).

Common SSIM to MOS mapping functions

Polynomial	Exponential	Logistic
$p_1x + p_0$	$a_1e^{b_1x}$	$100/[1+e^{-l_1(x-l_2)}]$
$p_2x^2 + p_1x + p_0$	$a_1e^{b_1x} + a_2e^{b_2x}$	

Parametric MOS models

We are using a combination of WR-model with SSIM

$$Q(D, \phi, u) = \alpha + \beta (1 + \gamma Q_{WR}(\phi, u)) f(D) + \delta Q_{WR}(\phi, u)$$

where:

- D – is the distortion measure (SSIM)
- $f(D)$ – mapping of SSIM to MOS scale
- ϕ – viewing angle of video as projected
- u – angular resolution of video
- $Q_{WR}(\phi, u)$ – Westerink-Roufs models
- $\alpha, \beta, \gamma, \delta$ – fitting constants

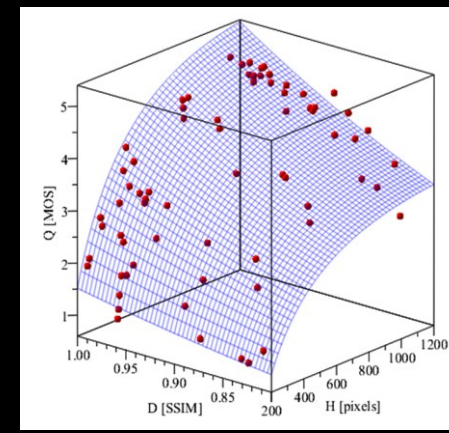
We use separate models for SDR and HDR cases.

Trained and validated on modern datasets.

References

- N. Barman, R. Vanam, Y. Reznik, "Parametric Quality Models for Multiscreen Video Systems," Proc. European Workshop on Visual Information Processing (EUVIP'22), Lisbon, Portugal, September 2022.
- N. Barman, R. Vanam, Y. Reznik, "Generalized Westerink-Roufs Model for Predicting Quality of Scaled Video," Proc. Int. Conf. on Quality of Multimedia Experience (QoMEX'22), Lippstadt, Germany, September 2022.

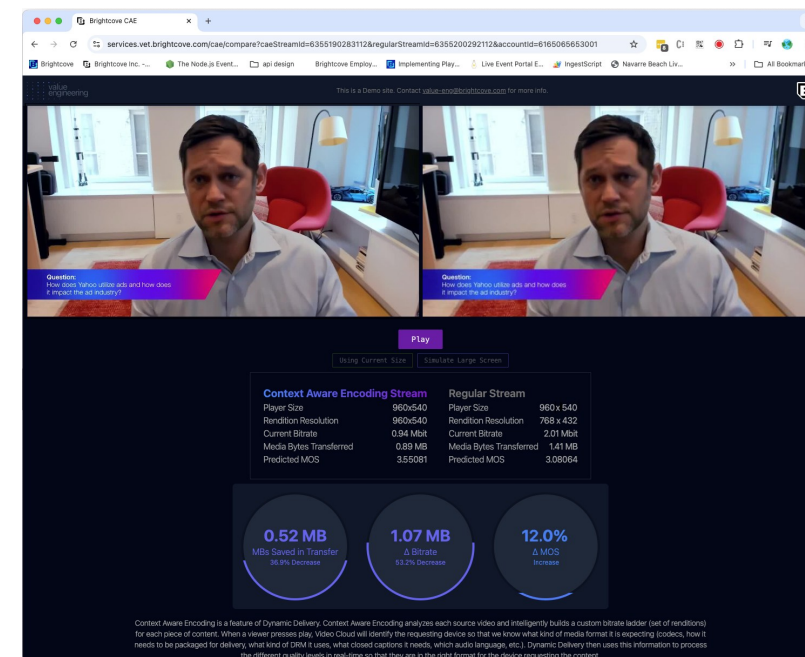
Model vs MOS scores in Netflix dataset



Example Implementation

Brightcove QOE analyzer tool

- Uses encoder-level signaling of SSIM scores
- Uses WR+SSIM2MOS model
- Reports predicted MOS scores specific to
 - Type of playback device
 - Display pixel density
 - Size of video player viewport
 - Resolution of encoded video
 - Statistics of SSIM scores in the content



Example Comparison

Brightcove CAE

services.vet.brightcove.com/cae/compare?caeStreamId=6355190283112®ularStreamId=6355200292112&accountId=6165065653001

Brightcove Brightcove Inc. ... The Node.js Event... api design Brightcove Employ... Implementing Play... Live Event Portal E... IngestScript Navarre Beach Liv... All Bookmarks

value engineering

This is a Demo site. Contact value-eng@brightcove.com for more info.

Question: How does Yahoo utilize ads and how does it impact the ad industry?

Play

Using Current Size Simulate Large Screen

Context Aware Encoding Stream		Regular Stream	
Player Size	960x540	Player Size	960 x 540
Rendition Resolution	960x540	Rendition Resolution	768 x 432
Current Bitrate	0.94 Mbit	Current Bitrate	2.01 Mbit
Media Bytes Transferred	0.89 MB	Media Bytes Transferred	1.41 MB
Predicted MOS	3.55081	Predicted MOS	3.08064

0.52 MB
MBs Saved in Transfer
36.9% Decrease

1.07 MB
Δ Bitrate
53.2% Decrease

12.0%
Δ MOS
Increase

Context Aware Encoding is a feature of Dynamic Delivery. Context Aware Encoding analyzes each source video and intelligently builds a custom bitrate ladder (set of renditions) for each piece of content. When a viewer presses play, Video Cloud will identify the requesting device so that we know what kind of media format it is expecting (codecs, how it needs to be packaged for delivery, what kind of DRM it uses, what closed captions it needs, which audio language, etc.). Dynamic Delivery then uses this information to process the different quality levels in real-time so that they are in the right format for the device requesting the content.

References

- Y. Reznik, "Achieving consistent quality of video delivery with MPEG-DASH," MPEG input document M25996, Stockholm, July 2012.
- Y. Reznik, "Proposed additional attributes for MPEG-DASH," MPEG input document M25997, MPEG-101, Stockholm, July 2012.
- S. Zhang, X. Wang, R. Yue, S. Cheng, A. Begen, and Y. Reznik, "Quality Information Carriage in File Format," MPEG input document M30350, MPEG105 - Vienna, Austria, July-August 2013.
- N. Barman, R. Vanam, Y. Reznik, "Parametric Quality Models for Multiscreen Video Systems," Proc. European Workshop on Visual Information Processing (EUVIP'22), Lisbon, Portugal, September 2022.
- N. Barman, R. Vanam, Y. Reznik, "Generalized Westerink-Roufs Model for Predicting Quality of Scaled Video," Proc. Int. Conf. on Quality of Multimedia Experience (QoMEX'22), Lippstadt, Germany, September 2022.
- J. Westerink and J. Roufs, "Subjective image quality as a function of viewing distance resolution and picture size," SMPTE Journal, vol. 98, 1989, pp. 113-19.
- Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, "Image quality assessment: from error visibility to structural similarity". IEEE Transactions on Image Processing 13 (4): 600–612, April, 2004.
- U. Pal and W. Law, Proposal for CMSD-based transmission of Media Quality Assessment (MQA) data, CTA WAVE, 27 Nov 2024

THANK YOU !