



## Case Study – Automated QA/QC Testing

*Bill Reckwerdt, CTO  
Video Clarity, Inc.*

Quality Assurance (QA) and Quality Control (QC) refer to systematically monitoring and evaluating various aspects of a project, service or facility to ensure that standards of quality are being met. It is important to note that quality is determined by the intended users and thus is not quantitatively definable.

Of course, if we set a goal and conduct subjective assessment, then we can create expectations. Various factors influence how we determine quality. Video quality is probably the most important but other factors include:

- 1) Synchronization; i.e. lip-sync.
- 2) The artistic content
- 3) The purpose of the viewer; i.e. watching a film versus a short clip on a mobile phone
- 4) Display type – LED TV, Plasma, LCD, or CRT
- 5) Viewing distance, ambient lighting, acoustic quality

This paper will explore how to set expectations and judge them in an automated way.

## Background

As service providers deliver multi-play, 3-screen offerings, performance measurements that provide insight into the customer's perception of quality is needed. Subjective quality tests are widely used to measure quality. However, the cost to conduct a statistically relevant assessment can be astronomical, and subjective quality testing is not practical for in-service, long duration, and repeatable monitoring.

Thus, the goal is to create an objective model to estimate the perceived quality of the audio and video. In other words if we want to use an oxymoron, the goal is to create an objective metric which measures subjective quality.

Objective models can use different approaches including:

- Full reference,
- No reference, and
- Reduced Reference.

Objective metrics may analyze

- individual pixels or samples,
- the bit stream, and
- the transmitted, packetized data.

### *When to use which*

In a lab or off-line environment, it is normally best to use a full-reference metric to judge quality. Full reference metrics perform a frame-by-frame comparison between the reference and the processed. This eliminates artistic content from having an effect on perceptual quality as you know what each frame should look and sound like.

Full reference has 2 drawbacks:

- The reference and processed sequences must be co-located
- The reference and processed sequences must be spatially and temporally aligned so every frame can be precisely matched.

In an in-service environment where the reference is not co-located, full reference is not feasible. Considerable research has been applied to no-reference techniques. These mainly look for blockiness and blur and make a quantitative assessment based on the amount of each.

No reference has 1 drawback:

- Artistic content can easily fool the algorithm – i.e. adding distortions or black frames to generate suspense will be generally marked as a quality drop.

Reduced reference metrics are a compromise between the full reference and no reference. They extract a number of features from the reference and compare the two videos based on just the extracted features.

Reduced reference has 2 drawbacks

- The reference and processed meta-data must be co-located. While this is not as much data as the full A/V sequence, it does mean that a back-channel is necessary.
- The reference and processed meta-data must be aligned. Once again this is not as complicated as full-reference, but it is an important consideration.

Remember that we rarely watch video without sound so the use of video throughout this paper implies audio – i.e. all of the metrics must analyze both audio and video perceived quality.

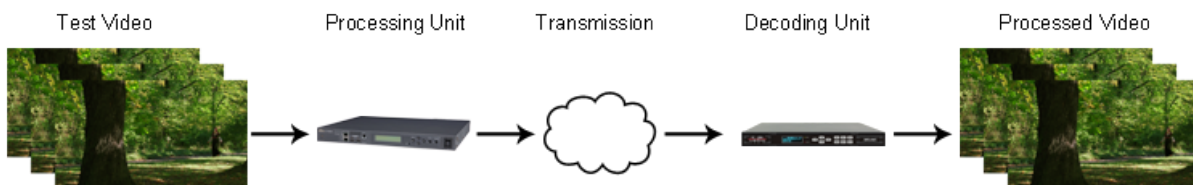
Moreover, one of the biggest items that affects quality is the synchronization between audio and video – i.e. lip-sync. If the video is ahead of the audio, then a human viewer can tolerate synchronization delay up to 125ms; however, if the audio is ahead of the video, then drops down to 45ms – in the real world light travels faster than sound so we are used to sound being behind.

## General Cast Study Flow

For this case study, we are setting up an automated test in a lab environment. Access to the reference and processed sequences is assumed so we will be utilizing full reference metrics.

A simplified drawing of a long duration test setup could look like this.

Figure 1: General Flow



Processing Steps:

- 1) Content is prepared and designated as a “test video.”
- 2) It is sent into a Processing Unit which can be as simple as compression device, but might include scalars, format converters, logo inserters, etc.
- 3) The processed video is then sent over a transmission medium (IP, Cable, Microwave, Satellite)
- 4) A receiver decodes the transmitted data and prepares it for display (Set-top box, phone, TV, PC)

### *Test Video*

Any test pattern can be defined. Normally, it is prudent to choose one with considerable temporal motion (frame-to-frame changes) and/or with spatial activity (forests, wind, smoke).

Several test sequences are defined by the EBU, VQEG, NTT, and LIVE (University of Texas – Austin). These are generally free to use, and fully annotated – the environment and original quality is defined.

For this test, we used the Park Joy HD video test clip from the EBU and combined it with a series of piano notes.

### *Processing Unit*

Video processing is a special case of general signal-processing where the input and output signal is a video stream or file. The most common form of video processing is to compress the video to fit it into the available bandwidth. The device that does this is termed a video encoder.

A video encoder reduces the quantity of data by removing the redundant information within a frame and between two successive frames. The digital video formats defined by VCEQ and MPEG are the de-facto standard for broadcast video. They are popular because

- There are no restrictions on the implementation of the video encoder (compression device).
- The video decoder's (Set-top box, PC) capabilities are fully-defined based on levels and profiles.
- The standards include video, audio, transport, and timing functions.

These video formats include - MPEG-1, MPEG-2 (DVD), H.263 (video surveillance), MPEG-4/H.264 (combined next generation standard), JPEG (still pictures), JPEG-2000 (archival) – just to name a few. With the possible exception of JPEG-2000, all of them are lossy (information is lost during compression so the quality after encoding/decoding is not as good as the original). JPEG-2000 can be lossy or mathematically loss-less.

In practice all lossy video encoders generate artifacts (areas of unfaithful visual/audible reproduction). If the encoder is designed and configured well and the data rate is high enough, then these artifacts will be virtually invisible. But video encoders do not work in a vacuum. Other factors such as transmission systems, temperature, and bad inputs can cause errors even to the best designed video encoder.

For this test, we use Adobe Premiere with the Mainconcept plug-ins to create MPEG-2 video and MPEG-1 Layer 2 audio processed test clips. Using the editor, we inserted some blank frames and caused some visual defects as noted later.

### *Video Transmission*

Video is transferred over a point-to-point, point-to-multipoint transmission medium. Examples include copper wires, optical fibers, wireless, or storage (DVDs, flash drives, and hard disks).

Even in guaranteed service networks, bit errors do occur. The streams are sent over many routers and any one of them can delay the packets (causing jitter), reroute the packets (causing loss or reordering), or simply fail.

One of the easiest ways to create an error is to oversubscribe (overbook) the network. This is the process of connecting more video streams to the network than can be supported if all of them operate

at peak usage. In normal operations, the network never runs at peak usage so this condition is not fulfilled. If it is, then packets are dropped, which will cause the video to stop or skip.

For this test, we did not use a network. We used a file-based work-flow.

### *Video Decoder*

Video decoders fall into 3 categories:

- 1) Professional grade integrated receiver/decoder (IRD). These normally have professional I/O connections like HD-SDI, SDI-3G, DVB-ASI, etc.
- 2) Consumer-grade converter/cable boxes called set-top boxes (STB). They have consumer I/O connections like component, SCART, HDMI, etc.
- 3) Software decoders built into a PC, tablet, or mobile device.

Video decoders are computerized devices, which receive compressed digital signals, decrypt/decode them, and convert them to either an analog or digital format to be shown on a TV. The video decoder can be either an external box, built into the TV, a PC, a gaming console, etc. Regardless, it makes it possible to receive and display TV signals, connect to networks, play games, and surf the Internet. One of its primary functions is to detect errors, and fix or conceal them. It does this by:

- 1) Holding previous frame/partial picture
- 2) Asking for a retransmission (Microsoft's IPTV solution)

For this test, we used the software decoders within our [ClearView A/V Analysis system](#).

### **Test Setup**

ParkJoy was ingested into Adobe Premiere along with 2-channel (stereo) audio (piano notes). Using the editor, a disturbance was added to 5 of the first 6 frames of the video in one test case and the video was blanked and the audio was silenced in other test case.

The two test cases were compressed using MPEG-2 video and MPEG-1 Layer 2 audio – both using the best quality settings at the chosen bit rate with no pre-processing.

All of these were ingested into ClearView using a simple server/client-based script as shown below.

```
REM Import ParkJoy Reference
cv LibraryActivate "G:\Test"
cv import "F:\ParkJoy_25.mov" ParkJoy_25 0 -1 -1
REM Import ParkJoy Processed Sequences
cv LibraryActivate "H:\Test"
cv import "F:\ParkJoy_25_Holes_15MP2_384MP1.m2t" ParkJoy_25_Holes_15MP2_384MP1 0 -1 -1
cv import "F:\ParkJoy_25_20MP2_192MP1.m2t" ParkJoy_25_20MP2_192MP1 0 -1 -1
```

Video Clarity offers two types of automated tests at this point:

- RTM – for long duration QA monitoring
- ClearView – for detailed perceptual analysis

### *RTM*

What would happen if the video processing units did not produce an error for several hours or days? Perhaps a particular set of input data sent at just the wrong time was needed to create the problem. This type of problem is very difficult to replicate, but it will be the first problem that your customer's find.

RTM can operate in the following 3 ways:

- 1) Compare 2 live inputs,
- 2) Compare a live input to a store sequence, and
- 3) Compare 2 stored sequences.

Regardless of the input, RTM continually monitors and records the A/V stream when the

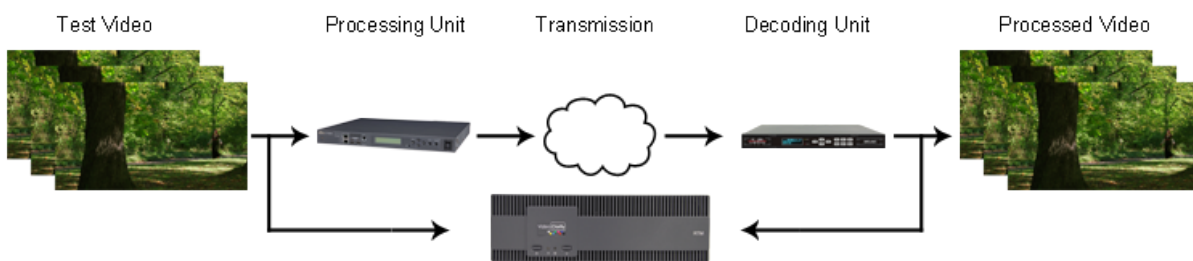
- Audio or Video quality drops below a defined threshold,
- Lip-sync exceeds the delay thresholds, or
- Ancillary data (VANC) is missing.

The recorded stream is stored in the ClearView sequence folder format for further analysis and classification.

In addition, RTM reports

- 1) The average A/V quality,
- 2) A/V delay/offset, and
- 3) Any dropped frames and then dynamically realigns.

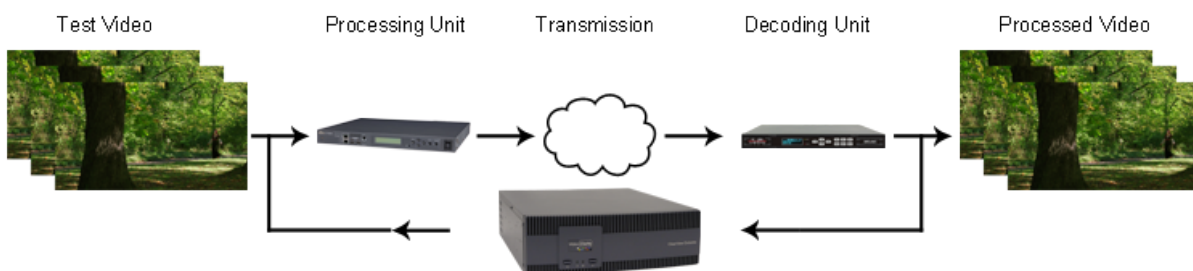
**Figure 2: Flow with RTM 3RU as the Dual Recorder**



### ClearView

For shorter duration testing (less than 3 hours), it is best to use a test & measurement video analyzer that can play uncompressed streams while simultaneously recording them.

**Figure 3: Flow with ClearView Extreme Broadcast as the Hardware Player & Recorder**



After the stream is recorded, ClearView aligns, processes, and reports A/V quality scores.

ClearView includes

- Preloaded VQEG & EBU Test Sequences,
- A [Multi-Format File Decoder](#),
- Automated spatial and temporal alignment routines,
- Perceptual Audio & Video Metrics (JND & DMOS),
- Objective Audio & Video Metrics (PSNR and Amplitude/Frequency),

- Side-by-Side [Viewing Modes](#), and
- GUI controlled or command line interface (CLI) with examples for python, perl, etc.

Since our test was a little over 20 seconds, we used ClearView.

We ran a simple script. For video, it calculates

- DMOS,
- JND, and
- PSNR.

For audio, it calculates DMOS

The script looks like this

```
REM DMOS,JND,PSNR,aDMOS Calculations ParkJoy versus 20Mbps MPEG-2
cv LibraryActivate "G:\Test"
cv mapA ParkJoy_25 -1 -1
cv LibraryActivate "H:\Test"
cv mapB ParkJoy_25_20MP2_192MP1 -1 -1
cv viewmode side
cv MetricWindow 0 0 1920 1080
cv SpatialAlign
cv stop
cv first
cv dmos "C:\Documents and Settings\user\Desktop\ParkJoy_25_20MP2_192MP1.dmos" -1 1 0
cv stop
cv first
cv jnd "C:\Documents and Settings\user\Desktop\ParkJoy_25_20MP2_192MP1.jnd" -1 -1 1 0
cv stop
cv first
cv psnr "C:\Documents and Settings\user\Desktop\ParkJoy_25_20MP2_192MP1.psnr" -1 -1 -1 0 1 0
cv stop
cv first
cv AudioMetricDMOS "C:\Documents and Settings\user\Desktop\ParkJoy_25_20MP2_192MP1.admos" two two
1 0 1
REM DMOS Calculations ParkJoy versus 15Mbps MPEG-2 with Holes in Audio/Video
cv LibraryActivate "H:\Test"
cv mapB ParkJoy_25_Holes_15MP2_384MP1 -1 -1
cv viewmode side
cv MetricWindow 0 0 1920 1080
cv SpatialAlign
cv stop
cv first
cv dmos "C:\Documents and Settings\user\Desktop\ParkJoy_25_Holes_15MP2_384MP1.dmos" -1 1 0
cv stop
cv first
cv jnd "C:\Documents and Settings\user\Desktop\ParkJoy_25_Holes_15MP2_384MP1.jnd" -1 -1 1 0
cv stop
cv first
cv psnr "C:\Documents and Settings\user\Desktop\ParkJoy_25_Holes_15MP2_384MP1.psnr" -1 -1 -1 0 1 0
cv stop
cv first
cv AudioMetricDMOS "C:\Documents and Settings\user\Desktop\ParkJoy_25_Holes_15MP2_384MP1.admos"
two two 1 0 1
.admos" two two 1 0 1
.dmos" -1 1 0
cv stop
```

The results graphed over time are shown below using the DMOS scale. Zero is defined as perfect; while 4 is of poor quality.

Our Video DMOS is calculated per field since the data is interlaced.

Our Audio DMOS is a 5 second sliding window in ½ second increments with each channel calculated separately.

The audio/video synchronization was calculated as 0.13ms.

For ParkJoy\_25\_20MP2\_192MP1,

- The content is 1080i/50Hz,
- Video compressed at 20Mbps, MPEG-2,
- Audio compressed at 192Kbps, MPEG-1 Layer 2, and
- Video frames 2-6 were artificially destroyed to create an affect. It was pretty hard to see subjectively as it was only visible for 0.2 seconds, but the objective metric found it easily.

**Figure 4: Park Joy first 4 Frames**



Figure 5: Video DMOS (measured in video fields)

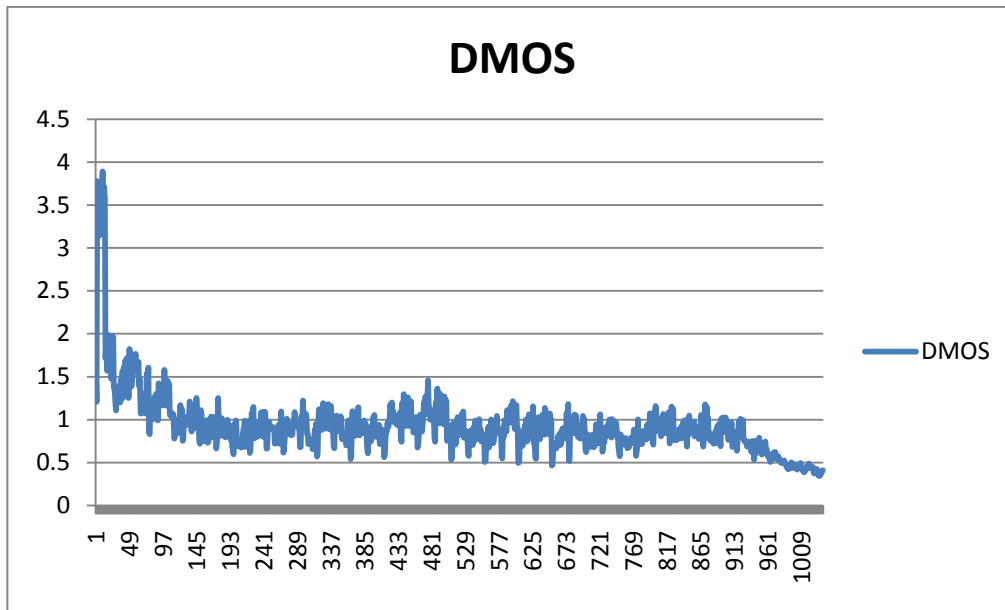
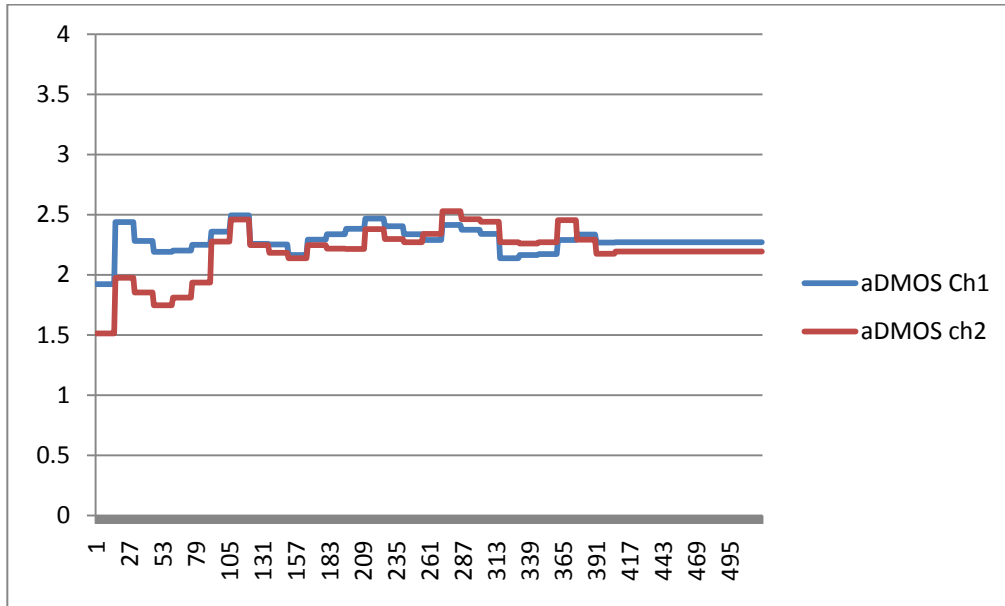


Figure 6: Audio DMOS (measured in video frames)



For ParkJoy\_25\_Holes\_15MP2\_384MP1,

- The content is 1080i/50Hz,
- Video compressed at 15Mbps, MPEG-2,
- Audio compressed at 384Kbps, MPEG-1 Layer 2, and
- The video was blanked (black frames were inserted) between 8.0 to 8.5 seconds.
- The A/V was blanked between 12.0 to 12.5 seconds.

Figure 7: Video DMOS (measured in video fields)

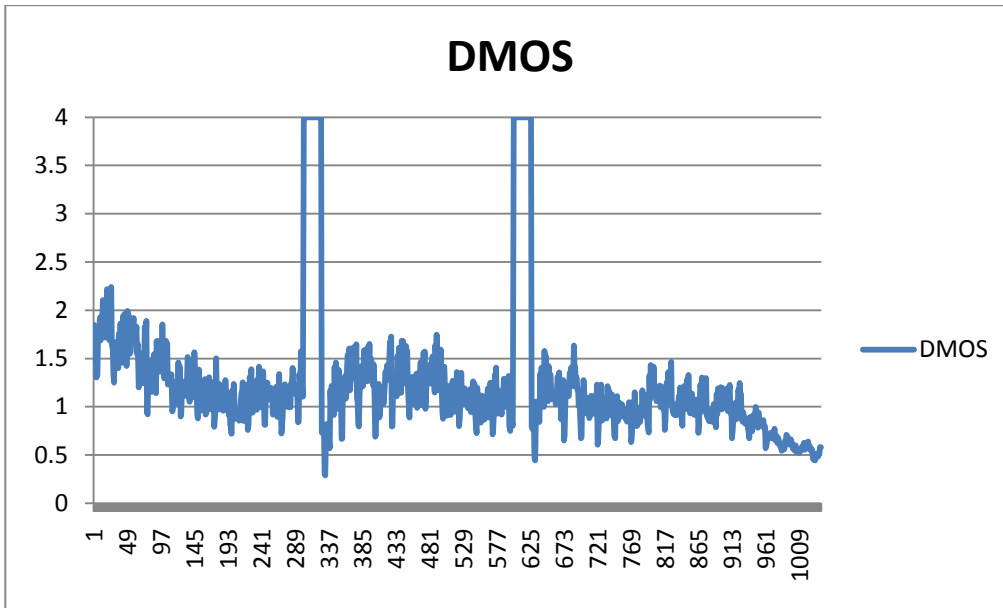
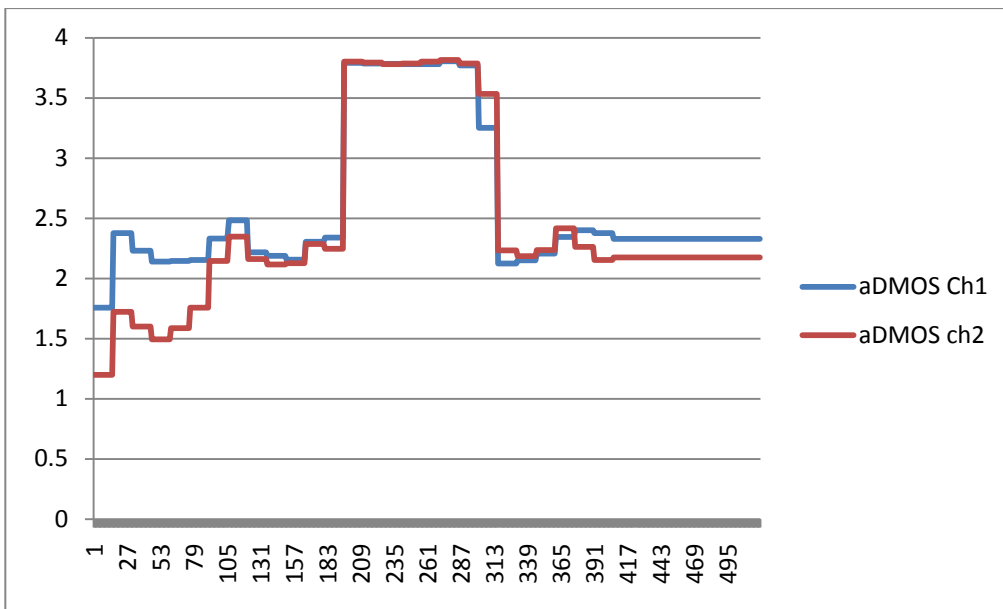


Figure 8: Audio DMOS (measured in video frames)



## Conclusion

We showed that our Objective Metrics correctly correlated with the predicted results of our test cases, and that they could be automatically run in a repeatable, scriptable way without human intervention.

## RTM and ClearView Availability

RTM and ClearView are currently being used by many broadcast equipment manufacturers. To get a demonstration, please contact Video Clarity or one of its channel partners ([www.videoclarity.com/sales.html](http://www.videoclarity.com/sales.html)) or visit us at one of the shows listed at [www.videoclarity.com](http://www.videoclarity.com).

## The Author

Bill Reckwerdt has been involved in digital video since the early 90's from digital compression, video on demand, to streaming servers. He received his MS specializing in Behavioral Modeling and Design Automation from the University of Illinois Urbana-Champaign.

He is currently the VP of Marketing and the CTO for Video Clarity, which makes quantitative, repeatable video quality testing and monitoring tools.