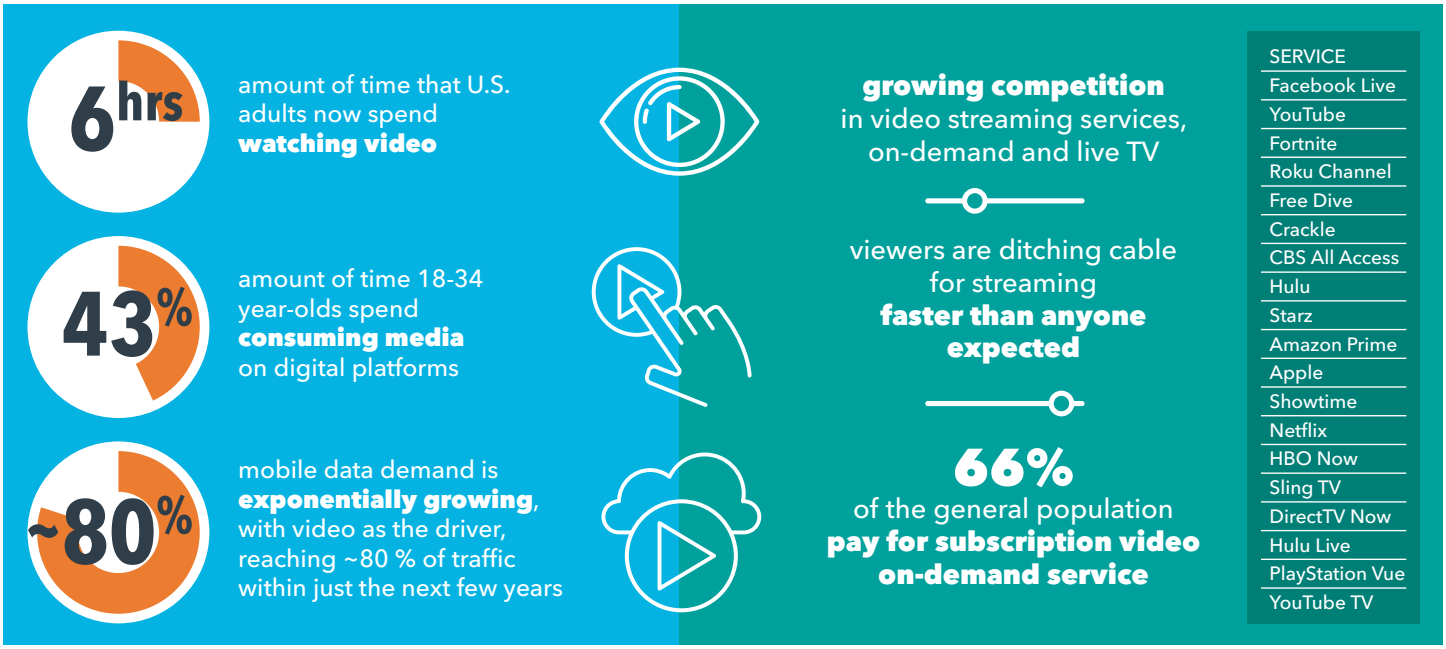
White Paper

Using AI to Assure the Viewer Experience for Streaming Video

Protecting Your Video Content Investment

Today, in the US, we spend 6 hours per day watching video, with all our video-capable devices and all types of video and streaming services. The younger generation is watching 43% of the time on non-TV media like tablets, smartphones and PCs, and we now have a vast selection of streaming services. 66% of the general US population has a subscription for a 'Video on Demand' service, and, as we consume video over multiple devices and platforms, the expectation is that video quality should be the same as broadcast.

As a result, close to 80% of all traffic on the mobile network is video. From the cellular side, US smartphone penetration is approaching saturation, causing mobile carriers to fight over each other's subscribers. About 95% of Americans own a cell phone, and 77% own a smartphone, making new subscribers a rare commodity. Hence, keeping and upselling to existing customers becomes more critical than ever.



"If someone is paying a few dollars a month for something, they're willing to put up with a few hang-ups now and again," says Brett Sappington, Parks Associates senior director of research. "It's also related to how often they use it. If they use an app daily and run into this problem daily, that hits the perceived value pretty hard."

Live content such as sports and newscasts is particularly challenging when it comes to measuring QoE. "Live is not easy," Sappington says. "That's kind of the last frontier to be conquered by the over-the-top (OTT) technology people. That's really the one that provides fits for delivery companies: how you deal with live, particularly live with high volumes of viewing."

This paper discusses how providers, network operators and manufacturers of smart TVs, streaming players, residential gateways and mobile devices can assess their IP video services' QoE. It's geared toward incumbent vMVPD and SVOD providers, as well as new and prospective market entrants such as mobile operators developing home video services for delivery over 5G. The paper explores topics such as:

- Why the entire streaming video ecosystem needs a **controllable, reliable and repeatable method** for evaluating and assuring video quality.
- The four primary methods for measuring video quality.
- Why the combination of **artificial intelligence (AI) and a unique non-reference algorithm** is the best way to understand QoE for live content, such as sports, news, and live events.

What is this Streaming Viewer Experiencing?

A. Freezing?
B. Packet loss?
C. Impaired frames?
D. All of the above?

If you're a virtual multichannel video programming distributor (vMVPD) or a subscription video on demand (SVOD) provider, you might choose D. And you could be right.

But if you're the customer of an vMVPD or SVOD, the right answer could be C—as in churn. With no contracts and several streaming providers to choose from, it's free and easy to switch when your current service keeps freezing, tiling and buffering like internet video circa 1996.

As the streaming IP video market matures, so do customer expectations about the **quality of experience** (QoE) they get from those services. Those expectations are set partly by their experiences with linear and on-demand services from traditional cable, satellite and telco TV providers. They're also set by how much they pay for a streaming service and how often they use it.

Typical Problems with Video

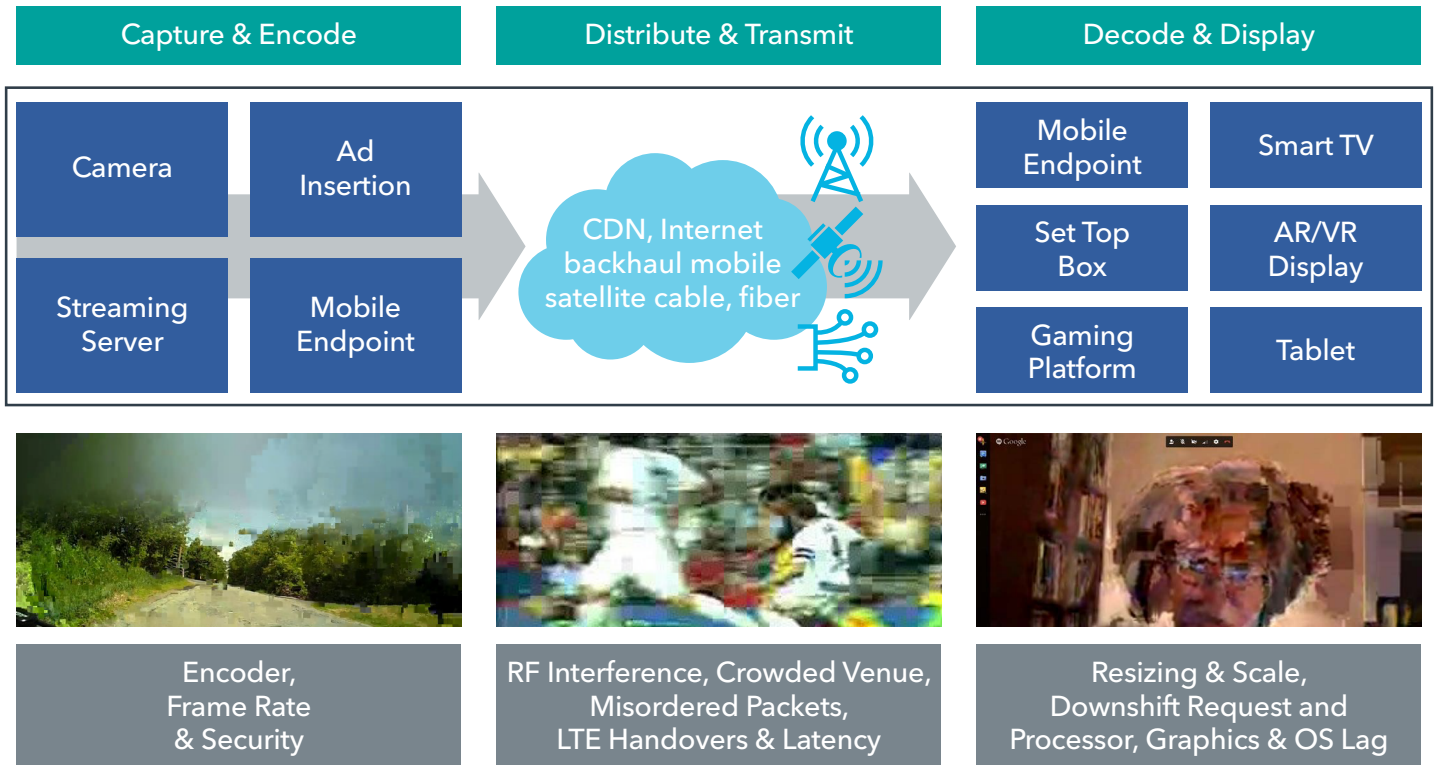
Video impairments fall within one of three categories:

Network Impairments	Encoding Impairments	Camera Impairments
Freezing	Blockiness	Noise
Buffering	Blurriness	Focus
Packet loss	Scaling	

Four Methods for Measuring Video Quality

The customer experience is where video quality problems come to the fore, and it's not unusual for customers to try to ferret out the problem. For example, if their DIRECTV NOW video keeps freezing or tiling, they might launch their Netflix, HBO or Hulu app to see whether those are also having problems. If so, they might deduce that their home network or their ISP is the problem. "For a consumer—and sometimes for the services themselves—it's really hard to pinpoint exactly where the problem rests," Sappington says.

So when a customer reports a problem, the provider needs to be able to determine from where it's originating. Figure 1 shows all of the primary points where artifacts can be introduced.



Artifacts can be introduced at many points along the video distribution and reproduction chain: at the capture, during encoding, while traversing the network, and during the decode and display.

This section discusses the four primary options that streaming providers, their network operator partners and endpoint vendors have for measuring video quality:

1. Measure packets
2. Compare frames
3. Use an algorithm to compare pixels
4. View pixels like a human

1 Measure Packets

This approach analyzes the delivery of the data that makes up the video stream. The value of its insights varies by content type. For example, packet drops and delays have a bigger impact on the QoE of live streams such as sports than they do with on-demand content such as movies. Also, virtually all vMVPD and SVOD content is encrypted. That rules out deep packet inspection (DPI), which is the only way to understand how much those drops and delays are undermining QoE.

The bottom line is that although measuring packets provides some useful insights, it's a coarse analysis that needs to be used in conjunction with other methods, such as signal strength, to get a fuller picture of the customer's QoE. For example, the packets might be delivered without a hitch, only to be mishandled by the endpoint's hardware or software. "You may not necessarily figure out what they see," Sappington says. "The processing and display of the video may be messed up."

2 Compare Frames

Comparing frames involves capturing video at the customer endpoint and then comparing it with a pristine reference. The approach makes it easy to spot problems that viewers would notice. Examples include delayed or missing frames, stalls and freezes, as well as audio and video that are out of sync. But this method requires unencoded video and the ability to mark the frames, both of which are impractical with vMVPD and SVOD services.

3 Use an Algorithm to Compare Pixels

This approach is similar to the previous one because it compares the received video to the source. This comparison uses full reference algorithms such as **Peak Signal-to-Noise Ratio** (PSNR), which roughly takes the RGB value difference for every pixel and uses that to compute a score. Another, more sophisticated algorithm is **Perceptual Evaluation of Video Quality** (PEVQ) which uses models that replicate human vision to evaluate blockiness, blur, noise and other attributes.

PEVQ is the better choice for creating a video mean opinion score (VMOS) that correlates with how real people score videos. But the drawback of PEVQ and other similar algorithms is that the test requires access to the source video.

4 View Pixels Like a Human

The new standard of video QoE analysis is a **non-reference algorithm that can look at the video just like a human would** and then give it a VMOS that tightly correlates with what a human would see. As its name implies, a non-reference algorithm doesn't require access to the source video to make a comparison, thus making it practical for real-world use. Non-reference video testing is critical for live content because, unlike on-demand, you don't have a source available that you can use for comparison.

For example, a network provider could use this approach to evaluate the QoE of third-party streaming services on its network and then compare that to the QoE on rival providers. This capability is particularly valuable for providers that are using cellular or other wireless technologies, which are subject to more anomalies than copper or fiber. As a result, providers can benchmark their services against those of their wired and wireless competitors to understand how their own video services' QoE compares to the rest of the marketplace. Video app developers also benefit because they can use lab tests to see if changes to device hardware, firmware and software affect QoE.

AI Enables the Epitome of Video QoE Analysis

The **artificial intelligence** (AI) field has advanced to the point that it's now capable of providing the kind of automated **video quality analysis** (VQA) that content providers, network operators, endpoint vendors, systems designers and others have sought for decades. It takes a holistic, end-to-end view and enables a wide variety of testing scenarios, such as testing a prototype mobile phone or streaming player to analyze the video QoE it provides on multiple types of network technologies from multiple service providers. Another example is using AI VQA to ensure that new endpoint software releases or compression techniques won't undermine QoE. AI VQA is capable of quantifying the QoE impact no matter where the artifacts are introduced.

Compared to human scoring, such as those obtained in focus groups, **AI VQA is objective, faster, far more repeatable and less expensive**. AI VQA also enables the use of a non-reference algorithm and a subset of AI, called **machine learning**, which makes it possible to train the algorithm on thousands of diverse video samples. This process enables it to provide a VMOS based on de facto industry standards and correlated to human perceptual scoring. The scoring is accurate with all of the most widely used resolutions and frame rates, from 480p through 4K, and from 24 to 60 FPS.

How Non-Reference Algorithms are Trained

In an AI VQA that supports non-reference algorithms, the machine learning process feeds the algorithm hundreds of thousands of sample video clips—far more than any human could sit through. Each clip is accompanied by a video quality score based on a respected industry standard. This enables the algorithm to build up a working knowledge akin to "If I see this, then the score must be that." Once it's been fully trained, the algorithm is now capable of scoring video that it's never seen before, whether that's an on-demand movie or a live newscast. Hence the term "non-reference."

One of the best video scoring systems in the industry (and what Spirent uses) is Video Multimethod Assessment Fusion (VMAF), developed by Netflix and the University of Southern California. VMAF combines human perceptual vision modeling with AI, and has been shown to be superior to many other algorithms in terms of its ability to produce a score that is well-correlated to how humans rate video quality.

An AI VQA system uses several non-reference models, each of which is trained on a specific set of artifacts, such as compression and scaling, which are known to cause problems that customers can see. A variety of VQA models are available to serve as the foundation for a vendor's set of non-reference algorithms. One example is the Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) model, which is a natural-scene, statistic-based blind quality assessment tool developed at the University of Texas at Austin's Laboratory for Image and Video Engineering (LIVE). BRISQUE is now one of the best, most-used quality assessment tools in broadcast and content production environments, making it ideal for AI VQA systems designed for use with streaming IP video.

Verifying Accuracy in Identifying Artifacts

Once trained, each non-reference model is tested to see how its scoring compares to an industry standard-scored (in this case, VMAF) video that was not included as part of the initial training.

The Pearson correlation is then calculated between the intended score (via VMAF) and non-reference compression model score for each of the thousand compressed clips in the baseline data set. It achieves a correlation of over 90 percent, which indicates a high level of confidence that it provides an excellent method of scoring video content without using a reference for comparison. For more details on this, see Spirent's accompanying whitepaper, Measuring Video Quality Using AI – Why It's Relevant and Reliable.

As a non-reference type, the algorithm provides these and other insights without requiring access to the source video. This design makes it ideal for all types of streaming video (live and on-demand), all types of

providers (vMVPD and SVOD) and all types of networks (wired and wireless). This broad applicability also makes it ideal for video app developers and for manufacturers of endpoints such as smart TVs and players.

Assuring High-Quality Video Experiences

Although many streaming services got their start as low-cost, best-effort alternatives to traditional cable, satellite and telco TV, the market has quickly matured. Now customers expect live and on-demand services to provide a QoE just as good as, if not better than, traditional services.

To meet these expectations, the entire streaming ecosystem—from vMVPDs and SVODs to their network operator partners/resellers to endpoint vendors—needs a fundamentally new and better way of measuring QoE. Traditional approaches are based on assumptions that the video must look good because, for example, the packet error rate is below a certain threshold.

The combination of AI and non-reference algorithms provides **deeper, more actionable insights into the customer experience**. This approach to video testing allows organizations to analyze video in the same way a human would perceive it, providing the ability to focus on the actual user experience without the expense of having individuals physically scoring the video content. These insights can also be used for competitors' networks, enabling providers to benchmark themselves. At the rate the streaming market is maturing and growing, these new tools and insights are must-haves for any provider wanting to stay competitive in its markets and seeking assurance it will be able to deliver the QoE it promises to its customers.

Reducing Costs

A growing number of video providers, network operators and other companies are also looking to these AI-based QoE analytics tools for not only quality of experience assessment, but also as a means to help reduce network operating expense. One such example is in use at a major U.S. mobile virtual network operator (MVNO). As a wireless reseller, the MVNO was using video compression to reduce network usage, thus lower the associated fees paid to the mobile network operator. But at the same time, they were concerned with how this compression could adversely impact the mobile user experience. The MVNO now uses the tool to optimize the levels of compression to ensure that they won't undermine QoE.

Spirent's Non-Reference Solution

The Spirent Umetrix® Video solution can "view pixels like a person" and score QoE according to VMOS, as if hundreds of human viewers were watching and rating overall performance. Umetrix Video supports any video service (e.g., mobile, home, 5G applications) and analyzes the video content via Spirent's content-trained non-reference algorithm, which uses machine learning on thousands of sample videos to understand the variations in different types of content (sports, drama, animation). Content training is based on de facto industry standards that correlate to human perceptual scoring.

The result: faster and less expensive repeatable design validation, regression testing, and competitive benchmarking.

For more information, visit Spirent Umetrix.

About Spirent Communications

Spirent Communications (LSE: SPT) is a global leader with deep expertise and decades of experience in testing, assurance, analytics and security, serving developers, service providers, and enterprise networks.

We help bring clarity to increasingly complex technological and business challenges.

Spirent's customers have made a promise to their customers to deliver superior performance. Spirent assures that those promises are fulfilled.

For more information, visit: www.spirent.com

spirent™
Promise. Assured.

TEMPEST
TELECOM SOLUTIONS

Contact us today: 805.879.4800 | tempesttelecom.com