Begin with the End in Mind: A Unified End-to-End Quality-of-Experience Monitoring, Optimization and Management Framework

Zhou Wang, PhD

University of Waterloo, Canada, zhou.wang@uwaterloo.ca

Abdul Rehman, PhD

SSIMWAVE Inc., Canada, abdul.rehman@ssimwave.com

Written for presentation at the

SMPTE 2017 Annual Technical Conference & Exhibition

Abstract. *There has been an increasing consensus in the video distribution industry that the design and operation of the full video delivery chain needs to be driven by the quality-of-experience (QoE) appropriate to the end-user. In practice, however, this is an extremely difficult task, largely due to the lack of effective mechanisms to instantaneously measure or predict every viewer's QoE. Moreover, existing quality assurance methods operate independently at different points along the delivery chain, reporting partial, inopportune, and incoherent measurements. This siloed structure leads to a fragmented understanding of the resulting video as it moves through each stage of the delivery network - from the acquisition point and head-end down to the media data center, the network, and eventually the end-user's viewing devices.*

Here we propose a new framework that uses a unified end-to-end solution to produce consistent QoE scores at all points along the delivery chain under the same evaluation criterion. This is a framework that produces a clear picture instantaneously to operation engineers, managing

executives and content creators about how video QoE degrades along the chain, a framework that allows immediate issue identification, localization and resolution, a framework that enables quality and resource usage optimization of the delivery chain for each of its individual components or as a whole, and a framework that provides reliable predictive metrics for long-term strategic resource and infrastructure allocations.

The main challenge in the implementation of such a framework is to create a unified QoE metric or process by which to measure QoE that not only accurately predicts human QoE, but is also light-weight and versatile, readily plugged into multiple points in the video delivery chain as either a single-ended or double-ended measure, producing real-time QoE scores across a wide range of bitrates, video resolutions, frame rates and dynamic ranges, and combining presentation picture quality with video freezing and adaptive streaming events. We show that the SSIMPLUS metric offers the best promise to such a solution. Once such a framework and the QoE metric is deployed in a video distribution system, many benefits come naturally. We demonstrate the benefit using bandwidth optimization as an example.

Keywords. Quality-of-experience, video distribution system, video quality assessment, video streaming, end-to-end quality assessment, video encoding, video transcoding, adaptive streaming

Introduction

Video distribution services have been growing exponentially in the past decade, coinciding with the accelerated proliferation of video content and smart mobile devices [1]. The gigantic scale of video data transmission has been supported by a vast investment of money and resources, and has also helped major industrial players grow their revenue immensely. Nevertheless, typical consumers, while enjoying the video content delivered to their TVs, tablets, and smart phones, often complain about the quality of the video they are receiving and experiencing. A recent survey by Limelight Networks shows that 39% of video consumers are considering changing their providers in the next 12 months due to poor video quality [2]. Meanwhile, content producers and providers are concerned about whether their creative intent is being properly preserved during the video delivery process [3], [4]. Currently, major video distributors are still seeking effective and efficient approaches to manage consumers' quality of experience when visualizing the video content being delivered to them.

Quality assurance (QA) has long been recognized as an essential component to warrant the service of modern video distribution systems. Traditionally, QA has been network-centric, focusing on the quality-of-service (QoS) [5] provided to the users, where the key metrics are defined by the network service level parameters such as bitrate, package drop rate, and network delay, together with integrity checks that guarantee the videos to be properly played at user devices. While QoS metrics are useful for basic QA purposes, they have fundamental problems in tracking what the users are actually experiencing. For example, the same video stream displayed on two different types of user devices (e.g., TVs vs. smartphones) with different combinations of window sizes and pixel resolutions may lead to very different viewer experiences. Any freezing event on the users' devices could result in a strong negative impact on user experiences. Different perceptual artifacts produced by different video compression methods could produce annoying visual impairment. All of these are not accounted for by QoS measures. Consequently, Quality-of-Experience (QoE) [6], which measures "the overall acceptability of an application or service as perceived subjectively by the end-user" [7], has recently been set to replace the role of QoS.

Nevertheless, QoE performance measurement is a very difficult problem itself, and the practical usage of the term could vary significantly from one solution to another. In practice, some quick remedies are often mistakenly used to replace QoE measurement. For example, device playback behaviors such as statistics on the duration and frequency of video freezing events, may be employed to create a crude estimate of visual QoE. Such quick remedies only provide a very rough idea about how certain components of the video delivery system perform, but are far from what we need in terms of accuracy, comprehensiveness and versatility. Most importantly, the perceptual artifacts that affect picture quality are not properly measured, and the large perceptual differences due to viewing conditions, such as viewer device, viewing resolutions and frame rate are not taken into account properly. As a result, it becomes difficult to use such approaches to precisely localize quality problems, to recover from failures, to optimize system performance, and to manage the visual QoE of individual users.

Another common mistake is to equate bitrate with picture quality or QoE, and use bitrate to define the level of services. However, encoding two different videos with the same bitrate could result in a substantial difference in perceived picture quality, meaning bitrate and picture quality/QoE cannot be used interchangeably. This is in addition to the large differences in performance between different encoders/transcoders with different configurations. Even worse,

the actual user QoE varies depending on the device being used to display the video, another factor that cannot be taken into account by bitrate-driven video delivery strategies.

Here we propose to use a unified end-to-end framework to tackle the QoE monitoring, optimization and management problems as a whole. The principle of the framework is to "begin with the end in mind", because the visual QoE of end users determines the ultimate overall performance of a video delivery system. Keeping this principle in mind, any design and resource allocation in the video distribution system, regardless of if it is for an individual component at the head-end, media center, network, access server, user device, or the whole system, should be evaluated, compared and optimized for one criterion, i.e., the impact on end users' QoE. To make such a system work properly, the most challenging task is to devise a highly accurate, efficient and versatile QoE metric. Once such a metric is deployed throughout the video distribution system, many optimization solutions come into play naturally.

End-to-End Visual QoE Monitoring, Optimization and Management

A general framework of modern video distribution systems is shown in Fig. 1. When a video distributor receives a source video, it passes the video through a sophisticated video delivery chain consisting of a series of processing, encoding, transcoding, packaging, routing, streaming, decoding, and rendering stages before it is presented on the screen of individual users' viewing devices. As far as quality assurance is concerned, the user experience measured at the very end of the chain is what matters. However, only measuring QoE at the very end would not be sufficient to help localize problems that could occur at any point along the chain. Therefore, to ensure the video is faithfully and smoothly delivered to the consumer device, the ideal quality assurance method would be to have inspectors deployed at the very end and also at each of the transition points along the chain. Ideally, all of these inspectors would be humans, as illustrated on the top section of Fig. 1, so that any quality issue can be identified instantaneously. In practice, however, this is not feasible because it requires thousands of source video streams and millions of derivative streams (the actual scale of many real-world systems) to be evaluated continuously by human inspectors, who are a constrained and non-scalable resource that may behave inconsistently over time.

To overcome the problem with a viable solution, we propose to replace humans with objective QoE monitoring probes, which constantly predict human QoEs at the corresponding inspection spots, as shown in the middle section of Fig. 1. There are two essential properties of such QoE monitoring probes:

- First, The QoE probes "see" and "behave" like human inspectors. More specifically, they observe all the actual pixels of all video frames like humans, and they produce QoE scores just like what humans would say about the video quality when seeing the same video streams.

- Second, the QoE probes provide a "unified end-to-end" monitoring solution in the sense that the QoE evaluation methods at all transition points along the video delivery chain are designed under the same evaluation framework and compatible methodology to produce consistent quality scores that are directly comparable.
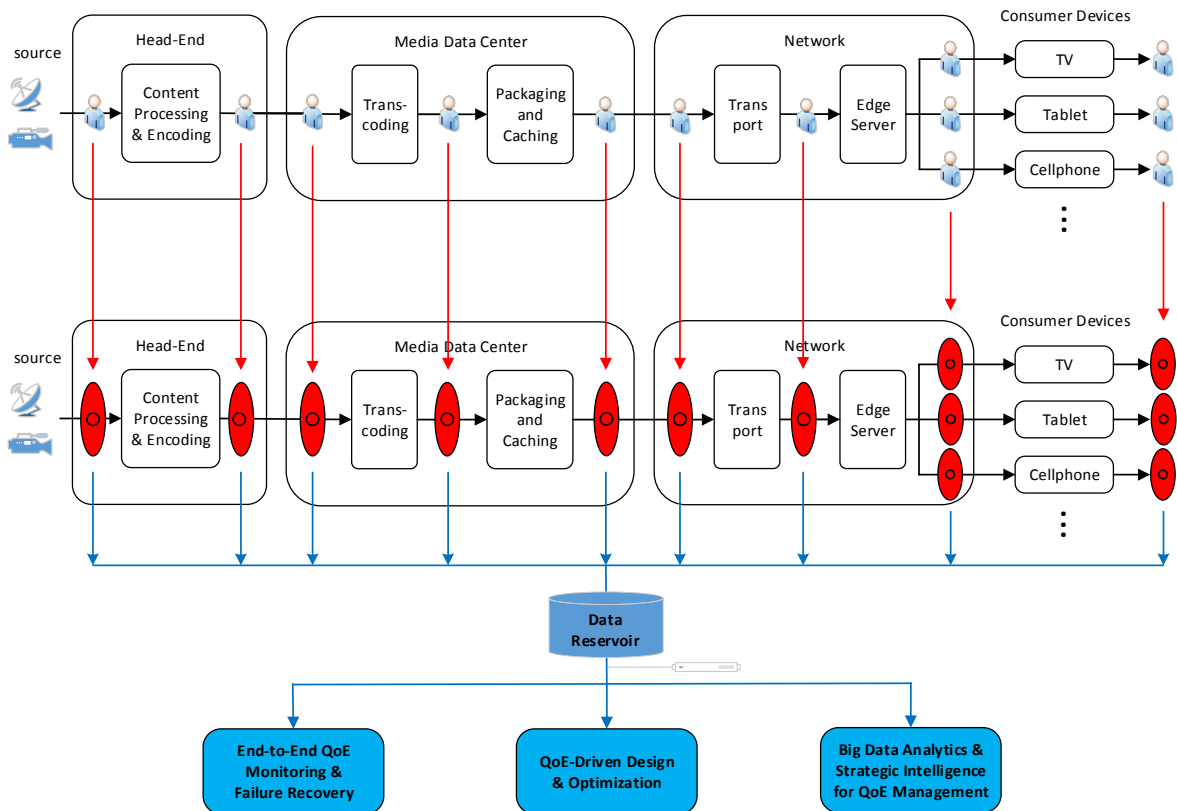
Figure 1. Unified end-to-end QoE monitoring, optimization and management framework in a video distribution system.

Having QoE monitoring probes deployed throughout the video delivery chain, QoE data is collected with statistics at different time-scales (minutes, hours, days, weeks, months, years), resulting in a valuable source for big data analytics and strategic intelligence.

Adopting the framework introduced above leads to many benefits.

- First, operation engineers will gain instantaneous awareness about how video QoE degrades along the chain, such that problems can be immediately identified, localized and resolved.

- Second, design engineers may closely observe the QoE of the input and output of individual components, and perform better design and optimization, and be confident about the impact of their new design and optimization on the final user QoE.

- Third, managing executives will have a clear picture about how video quality evolves throughout the video delivery system and over long time scales. Meanwhile, when long-time large-scale data has been collected, big data analytics can be performed, so as to make intelligent strategic decisions to manage user QoE.

Objective QoE Assessment Method

At the core of the end-to-end QoE monitoring framework is the QoE quality metric, which is also the most challenging technical problem to solve. Objective QoE prediction is difficult because it not only requires deep understanding of the human visual system, but also advanced computational models and algorithms, smart design and efficient implementation of the algorithms and systems. Traditional approaches such as peak signal-to-noise-ratio (PSNR) have been shown to have poor correlations with perceptual video quality. More advanced methods such as the structural similarity index (SSIM) [8], [9], multi-scale SSIM (MS-SSIM) [10], video quality model (VQM) [11] and video multi-method assessment fusion (VMAF) [12] significantly improve QoE predictions but are still limited in prediction accuracy. More importantly, these traditional approaches have fundamental limitations in their application scopes, functionalities and computational cost. These limitations largely impede these methods from being deployed broadly in real-world video distribution systems. The disadvantages of these traditional methods become even more pronounced when they are faced with the unified end-to-end QoE monitoring challenge we are targeting here.

An objective QoE metric that meets the challenge in a unified end-to-end QoE monitoring system requires the following features:

- ***Accurate and light-weight.*** The QoE metric must produce quality scores that accurately predict human visual QoE. The metric needs to be verified using independent, large-scale subject-rated video databases with diverse content and distortion types, and show high correlations with mean subjective opinions. Meanwhile, the metric needs to be light-weight (software implementation preferred), allowing for real-time computations of high resolution videos (e.g., UHD videos) with moderate hardware configurations. Such light-weight and speed requirement is critical in large-scale video distribution systems to reduce the overall cost and to maximize the flexibilities in terms of deployment, integration, customization, and scalability.

- ***Easy-to-understand and easy-to-use.*** The QoE metric must be easy-to-understand, directly producing QoE scores that linearly scale with what an average consumer would say about the video quality. For example, if the quality score range of the metric is between 0 and 100, then the total scale range may be divided into five even segments corresponding to five perceptual QoE categories of bad (0-19), poor (20-39), fair (40-59), good (60-79), and excellent (80-100), respectively. The QoE metric must be associated with a clearly defined hierarchical implementation structure and a carefully designed easy-to-use user interface (UI), where the main presentation is simple and intuitive, focusing on the most important trends and alert information. Such an easy-to-understand and easy-to-use QoE metric needs to define an easy-to-grasp language, under which executives will be able to make critical business decisions, and engineers will be able to identify and fix quality problems. An illustration of such a hierarchical structure and UI is given in Fig. 2, which provides a smooth transition between operational/tactical and business/strategic usage of the QoE metric.
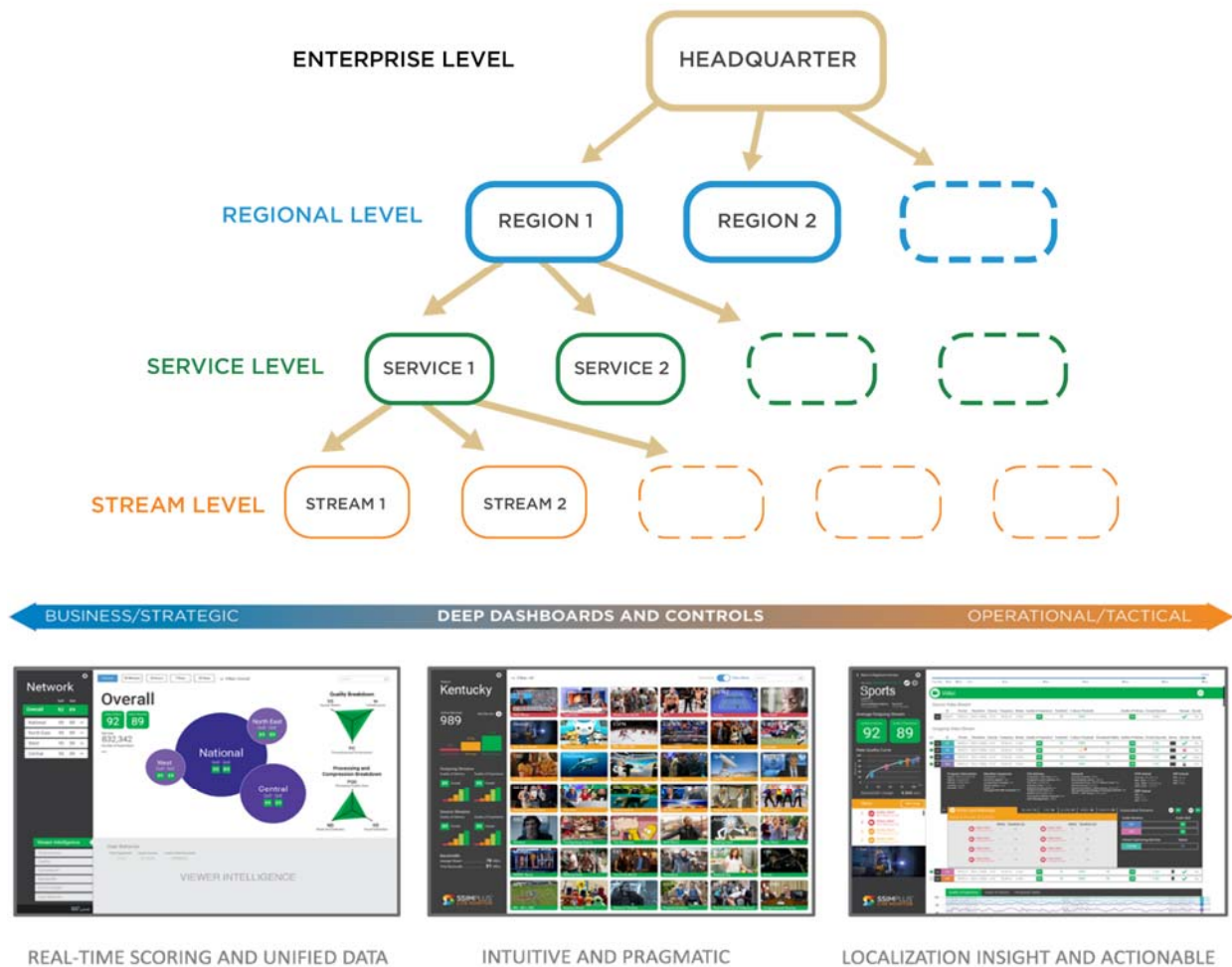
Figure 2. Hierarchical structure (top) and user interfaces at enterprise (bottom-left), region (bottom-middle) and service and stream (bottom-right) levels, that allow for a natural transition between operational/tactical and business/strategic usage of the QoE metric.

- ***Applicable and consistent across resolutions, frame rates, dynamic ranges, user devices and contents.*** In addition to accuracy and speed, another critical problem that hinders the wide usage of existing well-known video quality metrics (PSNR, SSIM, MS-SSIM, VQM, VMAF) is their limited applicability. In particular, when videos are of different spatial resolutions, frame rates, and dynamic ranges, these metrics are not applicable, because all of them require pixel-to-pixel correspondence. Moreover, when the same video stream is displayed on different viewing devices (e.g., TV vs. tablet vs. smartphone), the perceptual QoE could be significantly different. However, all of the metrics mentioned above only produce one quality score, and thus fail to make device-dependent QoE predictions. Furthermore, these quality metrics often produce inconsistent scores across different content types (e.g., sports vs. news vs. animations). As a result, when two videos with different content obtain similar quality scores, their perceived QoE may be very different. Such inconsistency strongly constrains the

usefulness of such QoE metrics in large-scale distribution systems that operate on thousands of video service channels to make resource allocation decisions across the whole systems. Therefore, to implement a unified end-to-end quality assurance framework for many real-world video distribution systems (e.g., for multi-screen and ABR video delivery networks), consistent and cross-resolution, cross-frame rate, cross-dynamic range, cross-viewing device, and cross-content QoE assessments are essential.

- ***Versatile for usage in single-ended, double-ended and more sophisticated scenarios.*** Single-ended and double-ended video quality assessments refer to the different application scenarios where a reference video may or may not be available when assessing the quality of a test video. Double-ended quality measures are essentially fidelity measures and PSNR, SSIM, MS-SSIM, VQM and VMAF all belong to this category, which assumes the reference video is accessible and of perfect quality. Double-ended quality measures typically have higher quality prediction accuracy than single-ended approaches, but are much more difficult to apply. Very often, the reference videos are completely inaccessible. Even when they are accessible, for example, at video transcoders, the reference videos are often not well aligned with the test videos both in space and time. Even worse, the source videos received from content providers are often distorted themselves, creating even more complex scenarios where the reference videos are already degraded. In order to provide consistent QoE assessment at all points along the video delivery chain, the QoE metric has to be extremely versatile. The QoE metric needs to be easily plugged into single-ended, double-ended and more sophisticated scenarios. It also needs to make the best use of all resources to produce the most accurate QoE prediction. For example, at the transcoder, the QoE metric needs to precisely align the source and test videos before applying double-ended fidelity assessment. It also needs to appropriately handle the case when the reference video quality is already degraded.

All of the above are must-have features for a QoE metric to work effectively in a unified end-to-end quality monitoring framework. Conventional and well-known video quality metrics (PSNR, SSIM, MS-SSIM, VQM, VMAF), however, are too far from meeting these requirements (and in most cases not applicable, regardless of accuracy and speed), and thus their usage is typically limited to lab-testing environments or restricted to small-scale use cases for testing certain individual components in the video delivery system.

The large gap between the limited performance and functionality of the well-known video quality metrics and the essential requirements of large-scale unified end-to-end QoE monitoring systems has motivated the development of the SSIMPLUS video QoE metric, which has been set to meet all the requirements throughout its design and implementation phases [13]. A recent study using 10 independent publicly-available subject-rated video databases (created from a collection of hundreds of thousands subjective ratings) evaluates conventional and state-of-the-art video quality metrics (including PSNR, SSIM, MS-SSIM, VMAF, SSIMPLUS and several other metrics), by comparing the quality predictions of these metrics against subjective mean opinion scores (MOS) [14]. The results showed that SSIMPLUS achieves the highest QoE prediction performance in terms of its correlation coefficients against MOS. It appears to be the only QoE metric that achieves an average correlation coefficient higher than 0.9. The same study also found that the SSIMPLUS metric to be 16.4 times faster than the VMAF metric, allowing SSIMPLUS to be computed in real-time in real-world applications [14]. The SSIMPLUS metric is applicable and produces consistent scores across resolutions, frame rates, dynamic ranges and content types. For every single video stream, it generates multiple QoE scores

corresponding to a wide spectrum of viewing devices, from small screens on cellphones to large-size TVs. When applied to ABR encoding, SSIMPLUS simultaneously computes single-ended QoE scores of the source video input, together with double-ended scores for all the derivative video output produced by transcoders with different bitrates and resolutions. As well, it provides the absolute QoE scores of the derivative streams considering that the source input does not have perfect quality. At the client side, SSIMPLUS combines picture presentation quality with the impact of switching and stalling events to produce an overall QoE assessment for each individual user on a per-view basis [15]. All of these computations are done at a speed faster than real-time. Due to these features, SSIMPLUS has been successfully deployed in large-scale operational environments, running 24/7 reliably.

QoE-Driven Bandwidth Optimization

Once a unified end-to-end QoE monitoring solution is in place, many benefits come as a natural next step, as described earlier. Many of such benefits involve optimization, and one widely recognized example is bandwidth optimization. Although a significant number of solutions have been proposed in the industry for saving bandwidth, talking about bandwidth reductions without maintaining the right level of visual QoE makes little sense. Due to the lack of proper QoE assessment tools, existing bandwidth saving approaches, whether it is applied to encoding/transcoding or streaming optimization, result in unstable results. To perform bandwidth optimization properly, the first step has to be adopting a trusted QoE metric with powerful functionalities, e.g., accurate, meaningful and consistent quality assessment cross resolutions, frame rates, dynamic ranges, viewing device and video content. Below we give an illustrative example using SSIMPLUS as the example QoE metric to demonstrate how large bandwidth savings can be achieved in live and file-based operations by making use of such a QoE metric.

Significant bandwidth savings can be obtained by adopting a QoE measure that produces consistent QoE assessment across content, resolution, and user device, each of which could lead to significant gain. A step-by-step example is given in Fig. 3. Firstly, because of the difference in encoding difficulty of different content (Title 1 and Title 2 shown in the top graph of Fig. 3), to reach a guaranteed QoE quality level (SSIMPLUS = 90), using a fixed bandwidth to encode all videos may be a waste, depending on video content, e.g., using a fixed 4Mbps when only 3.1Mbps is necessary for Title 2. Second, when the same content is encoded to two or more spatial/temporal resolutions, the capability of picking the most cost-effective spatial/temporal resolution to achieve the guaranteed quality level can also help save large bandwidth, e.g., a bandwidth reduction from 3.1Mbps to 2.4Mbps is obtained by switching from 1080p to 720p resolutions, as shown in the middle graph of Fig. 3. Finally, the perceptual QoE varies significantly on different viewing devices. This can be seen in the quality-bitrate curves in the bottom graph, which shows that when the user is known to use a smartphone rather than a TV to watch the video, a bandwidth of 0.8Mbps is sufficient to achieve the same target quality level (SSIMPLUS = 90). With all three factors combined, a total of 80% bandwidth savings may be obtained (from 4Mbps to 0.8Mbps).
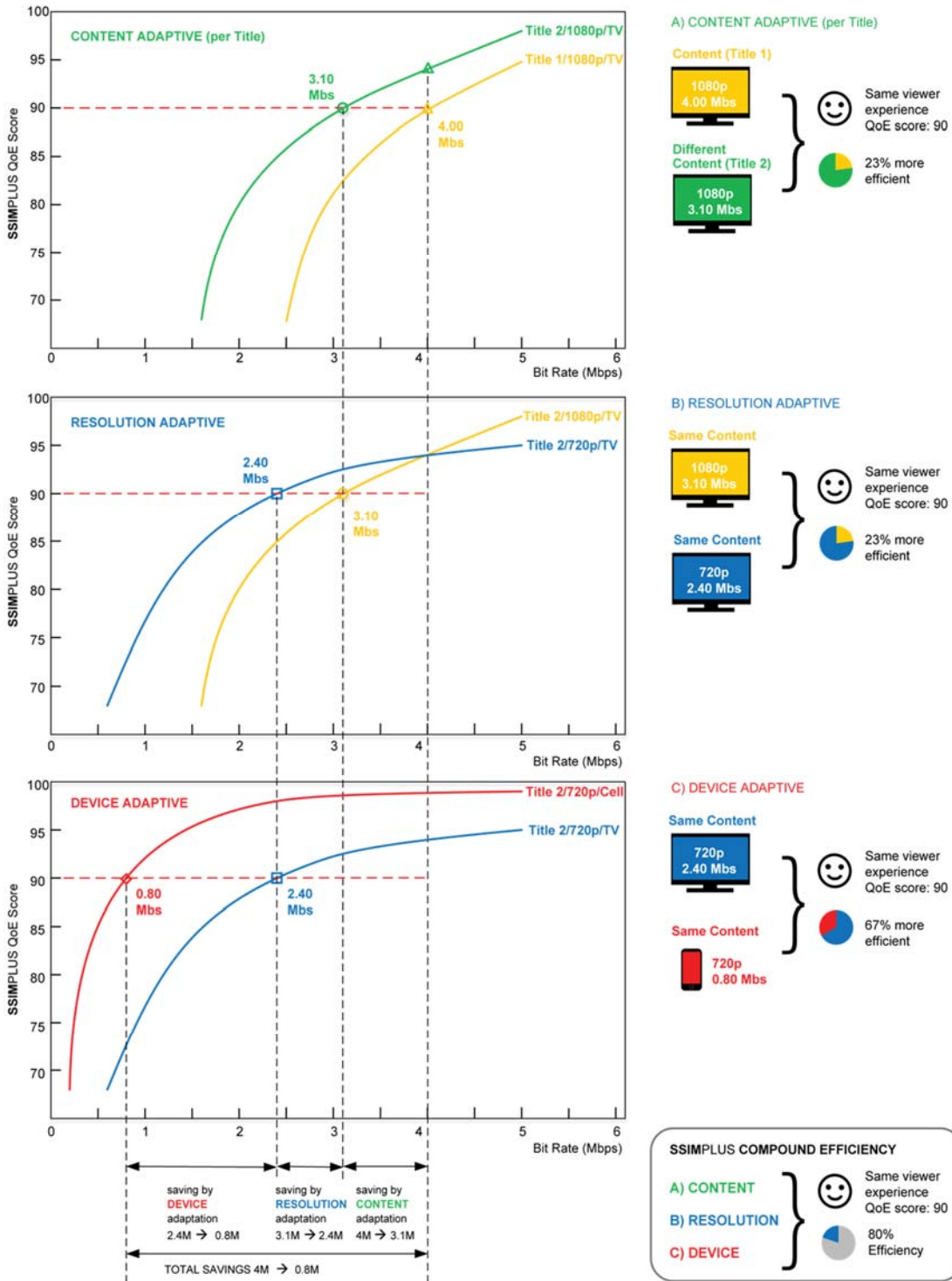
Figure. 3 Illustration of how bandwidth savings is achieved by using a QoE metric that is able to adapt to video content (top-graph), video resolution (middle-graph), and user viewing device (bottom-graph). All levels of adaptations contribute to the final bandwidth savings, without compromising the guaranteed user QoE level (SSIMPLUS = 90).

Although the example given here is for illustration purposes only, and in practice users may be constrained to explore all three factors for maximum cost-savings, our intensive study suggests that for most video content and the most common usage profiles, an average cost saving of 20%-60% is typically achieved by properly adopting this QoE metric-driven bandwidth optimization technology. Such bandwidth savings can be implemented by adaptive operation of video encoders/transcoders, and may also be incorporated into adaptive streaming frameworks to achieve similar goals in a dynamic way.

Conclusion

We propose a framework for unified end-to-end QoE monitoring, optimization and management. The principle behind the design of the framework is to start with end user's QoE. All the QoE monitoring points should produce instantaneous scoring that reflects the end user's QoE up to the monitoring point in the video delivery chain. The QoE scores need to be accurate, consistent and directly comparable, such that the monitoring solutions of the entire video distribution network speaks the same language, from the head-end, media center, network, access server, and client viewing device. Such a unified end-to-end solution laid the groundwork for the subsequent operations of great benefits. Specifically, operation engineers will be able to immediately identify, localize and resolve quality problem, design engineers will be able to perform effective and accurate optimizations on individual components on the video delivery chain, and managing executives will have a clear picture about how video quality evolves throughout the distribution system and over long time scales, so as to make intelligent strategic decisions to manage user QoE.

The most challenging task in implementing the proposed framework is to create an objective QoE metric that is not only accurate, fast, easy-to-understand and easy-to-use, but also applicable and consistent across resolutions, frame rates, dynamic ranges, viewer devices and contents. Moreover, it needs to be highly versatile for use in single-ended, double-ended and more sophisticated scenarios. Conventional and well-known video quality metrics such as PSNR, SSIM, MS-SSIM, VQM and VMAF are too far from meeting these requirements, in terms of not only accuracy and speed, but also applicability and functionality. As a result, their usage is limited to lab-testing environment or small-scale use cases. In today's environment, a metric is required that satisfies all critical requirements for a unified end-to-end QoE monitoring system. Such great needs have motivated the recent development of novel video QoE metrics such as SSIMPLUS [13], [14], which has been deployed in real-world large-scale QoE monitoring systems.

To further demonstrate the benefits of adopting the proposed framework and QoE metric, we use bandwidth optimization as an example, which demonstrates that large bandwidth savings can be obtained with little effort, purely by adopting a QoE metric such as SSIMPLUS.

With the wide deployment of the proposed framework and QoE metrics in large-scale video distribution networks. The QoE data collected in large and varying space and time-scales constitutes a valuable source for big data analytics and strategic intelligence which may be an interesting direction for future investigations.

References

1. Cisco Inc., "Cisco Visual Networking Index: Forecast and Methodology 2015-2020", 2016.

2. Limelight Networks, "The state of online video," https://www.limelight.com/video/, 2016.

3. C. Curtis, *et al.*, "American Society of Cinematographers Technology Committee Progress Report 2016," *SMPTE Motion Imaging Journal*, vol. 125, no. 7, pp. 43-58, Sept. 2016.

4. Z. Wang, "New quality-of-experience measurement technologies: streamlining how videos are delivered to consumers," *IEEE Signal Processing Society Blogs*, July 2017.

5. M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, Sept. 2014.

6. O. Oyman, S. Singh, "Quality of experience for HTTP adaptive streaming services," *IEEE Communications Magazine*, vol. 50, Apr. 2012.

7. ITU QoE Recommendation ITU-T P.10/G.100, Amd.1, *New Appendix I Definition of Quality of Experience (QoE)*, 2007.

8. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, Apr. 2004.

9. Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, Feb. 2004.

10. Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conference on Signals, Systems and Computers,* Nov. 2003.

11. M. H. Pinson, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312-322, Sept. 2004.

12. Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy and M. Manohara, *"Toward A Practical Perceptual Video Quality Metric*," *Netflix TechBlog, June, 2016*

13. A. Rehman, K. Zeng and Z. Wang, "Display device-adapted video quality-of-experience assessment," *IS&T/SPIE Electronic Imaging: Human Vision & Electronic Imaging*, Feb. 2015.

14. SSIMPLUS: The most accurate video quality measure, https://www.ssimwave.com/from-the-experts/ssimplus-the-most-accurate-video-quality-measure/

15. Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang "A quality-of-experience index for streaming video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 154-166, Feb. 2017.