QoE Assessment for Multi-Video Object Based Media

Tomasz Lyko*, Yehia Elkhatib†*, Michael Sparks‡‡, Nicholas Race* and Rajiv Ramdhany‡‡
*School of Computing and Communications, Lancaster University, United Kingdom, {i.lastname}@lancaster.ac.uk
†School of Computing Science, University of Glasgow, United Kingdom ‡‡BBC R&D North, United Kingdom

***Abstract*—Recent multimedia experiences using techniques such as DASH allow the streaming delivery to be adapted to suit network context. Object Based Media (OBM) provides even more flexibility as distinct media objects are streamed and combined based on user preferences, allowing the experience to be personalised for the user. As adaptation can lead to degradation, modelling and measuring Quality of Experience (QoE) are crucial to ensure a perceptibly-optimal user experience. QoE models proposed for DASH include quality-related factors from single video-object streams and hence, are unsuitable for multi-video OBM experiences. In this paper, we propose an objective method to quantify QoE for video-based OBM experiences. Our model provides different strategies to aggregate individual object QoE contributions for different OBM experience genres. We apply our model to a case study and contrast it with the QoE levels obtained using a standard QoE model for DASH.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Object Based Media (OBM) presents a new form of multimedia where the stream is identified by a number of separate objects that together form the overall media experience. The different objects are combined at the client, or at any point in the network, to create a personalised AV presentation for the user. OBM experiences can flex to adapt to user preferences, device types, and consumption patterns. By deferring the composition of the discrete media objects to occur downstream and in response to user input, OBM delivery platforms can create high degrees of personalisation and interactivity [1].

Research on Quality of Experience (QoE) for Dynamic Adaptive Streaming over HTTP (DASH) has involved development of Adaptive Bit Rate (ABR) algorithms as well as video quality metrics such as VMAF [2]. Whilst DASH can be used to deliver simple AV-only OBM experiences via its AdaptationSets (for different AV objects) and in-band event signalling (for object orchestration), typical applications use a combination of DASH streams and other assets for presentation. Though numerous QoE models have been proposed and standardised for DASH streams, they are focused exclusively on assessing video quality and playback experience. These models are not applicable to OBM as they do not cover the multi-faceted aspects of these experiences: multiple AV and non-AV objects, and their spatial and temporal relationships.

This leads us to revisit a fundamental research question: *how can we objectively assess the QoE of an OBM experience?* We approach this question by understanding the OBM context, identifying how a typical DASH QoE model operates for an OBM experience, and then developing a new model for OBM. Our contributions in this paper are:

- An overview of QoE methods for DASH (§II).
- An introduction to OBM (§III) and a description of the problem space (§IV).
- An analysis of the suitability of existing QoE approaches to OBM (§V-A), and a proposed OBM-specific QoE assessment model (§V-B).

II. Background

Here, we briefly introduce DASH, in which the content is encoded into multiple bitrates and divided into short and interchangeable segments of equal duration, all of which are described in a manifest file. The client selects an appropriate quality bitrate for each segment, aiming to maximise the QoE, which, can be divided into the following three main factors [3]: Encoding Quality [4]–[9], Quality Switching [7], [10]–[13] and Rebuffering [6], [14]–[18]. QoE models designed for adaptive streaming aim to provide a single metric that encompasses all or most of the aforementioned factors by quantifying their impact on the overall QoE of a streaming session [19]. There have been many models developed over the years [15], [17], [20]–[22] including one standardised model, ITU-T Rec P.1203 [23], that outputs a predicted Mean Opinion Score (MOS) on a 5-point scale designed for subjective testing as standardised in ITU-T Recommendation P.910 [24].

III. OBM Primer

OBM experiences place the user at the centre of the story by allowing the narrative and its presentation to be adapted to the user's preferences. Using spatial and temporal placement of media objects (audio, video and 2D/3D graphics) as well as multiple representations of these objects, they allow *(1)* the presentation to fit to user choices, device characteristics, and crucially, *(2)* the user to navigate alternative story pathways. OBM experiences can be consumed on heterogeneous devices (smart TVs, phones, VR) and networks. This renders tasks such as authoring, object-transformation, quality assessment and timely delivery challenging.

A. Multi-Video OBM Definition

OBM refers to a media-experience representation based on discrete atomised objects and a description of orchestration rules/constraints that govern their presentation. The description defines each object's semantic role in the story and organises all objects into editorially-coherent paths in a story graph. The assembly of media-objects is *dynamic* (responds to events such as user interaction), *context-sensitive* (uses user-data and device characteristics) and can happen anywhere in the distribution chain.

B. Examples

We created a number of trial experiences, testing various combinations of branching narratives; alternative presentation forms; optional display types, etc. ***BBC News Click 1000*** is a branching video narrative with just-in-time client decisions. **Forecaster** is a factual presentation that allows multiple renditions based on the audience, e.g. switching the presenter with a BSL signer, and hyperlocalisable overlays. is a factual programme that can be presented in either a documentary or entertainment-oriented track or a mixture of both, each with varying audio and video. This presents, based on user interests, location etc, a semantically-coherent collection of clips, sections and facts.

Delivering a coherent and smooth presentation is non-trivial; orchestration needs to account for/control characteristics of individual objects (e.g. encoding quality, presentation speed), their delivery (e.g. pre-fetching, concurrent decoding), and adjust inter-object factors (e.g. colour grading, audio mixing, playback synchronisation).

Given this, the notion of overall QoE poses challenges to the production process. There is clearly a relationship and balance between media-object-level presentation-quality metrics as well as experience-level ones. We also need to know what Encoding Qualities (e.g. resolutions) to generate for each object based on its user-perceived importance. These metrics can ensure that degradations are quantified and controlled (e.g. step up/down a provided Encoding Quality ladder) in an acceptable manner in a highly diverse delivery environment.

IV. Problem Space

We now comment on the suitability of DASH QoE models to OBM experiences, specifically in layered multi-video OBM.

A. Key differences between DASH and OBM

Linearity. An OBM experience can have more than one stream. This non-linear nature means that QoE measuring methods can vary during the lifetime of an OBM experience.
Differentiated priority. The relative importance of each substream to the overall 'experience' may be a factor that is determined by the production team on a 'per-show' basis.
Customisation. Dynamic customisation of elements is an inherent feature of OBM; e.g., giving substreams different priorities as reflected by their screen real estate properties.
Interactivity. This is the propensity of the user to customise the experience. In lean-forward OBM experiences (high user agency), the user actively interacts with the experience, e.g. selects accessibility options or story branches. In lean-back experiences (low user agency), customisation is based on pre-specified preferences at a much coarser grain.
Fluid resource utilisation. A mixture of resources could be used to deliver an OBM experience, ranging from cloud data-center resources, edge resources, network middleboxes, and end-user devices such as smart TVs, headsets, and set-top boxes. In this work, we focus only on client-side assembly.

B. Design goals

In light of these differences, QoE quantification techniques need to be *dynamic* and *customisable* to match the nature of OBM. Specifically, there is need for objective QoE models to:

- factor in the qualities of different elements,
- identify the importance of each element to the user,
- adjust to changes in the flow of the media content, timeliness in order to avoid *glitches*; different types of latency,
- adapt to the presentation context, specifically in terms of device capabilities and platform restrictions.

V. Proposed Model

When compared to typical adaptive streaming (e.g. DASH), OBM introduces additional complexity to the estimation and management of QoE in streaming sessions.

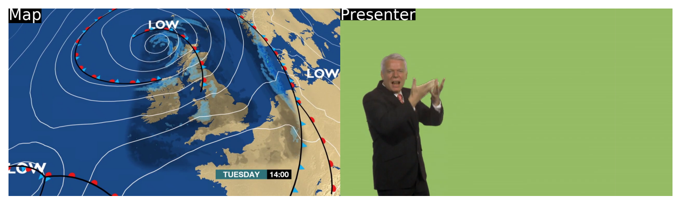

Fig. 1. A BBC weather forecast programme divided into two OBM objects.

We specify a case study to illustrate QoE assessment using DASH models. A typical weather forecast programme consists of the presenter with the map behind them, each of which can be delivered as a separate media object (Figure 1). The objects can be combined at the client or in the network to create a single video element. Each object can have several independent quality representations, resulting in a number of possible permutations of quality. Assuming a simple ABR ladder consisting of low, medium, and high quality representations corresponding to encoding settings of 416x234 at 145kbps, 768x432 at 730kbps and 1280x720 at 4500kbps respectively, based on the recommended specifications for Apple devices [25], there are nine possible quality permutations. In DASH, there would be only three quality permutations of the video.

A. QoE using P.1203

Using P.1203 [26]–[28] in the highest mode of operation, we can calculate the predicted MOS (1-5) of each quality permutation, as well as of the individual media objects. The input for the model for each quality permutation was created by combining both media objects together into a single video element, resembling the final video element that would be displayed at the client, since QoE models designed for adaptive streaming do not support multiple media objects. Media objects were encoded into H.264 using `x264` [29] and combined using `ffmpeg` [30]. As shown in Table I, we observe a significant disparity between the predicted scores of all quality representations for both media objects, with higher quality

TABLE I
QoE ASSESSMENT OF QUALITY PERMUTATIONS OF THE FORECASTER PROGRAMME USING MULTIPLE STRATEGIES AS WELL AS P.1203 ALONE, WITH MOS OF INDIVIDUAL OBJECTS SHOWN IN BRACKETS.

	Presenter_LQ (2.1)					**Presenter_MQ** (3.9)					**Presenter_HQ** (4.9)				
	P.1203	Avg	Size	SI	TI	P.1203	Avg	Size	SI	TI	P.1203	Avg	Size	SI	TI
Map_LQ (2.0)	2	2.1	2	2	2.1	3.5	3	2.3	2.3	3.2	4.6	3.4	2.5	2.4	3.9
Map_MQ (3.5)	3.5	2.8	3.3	3.3	2.6	3.5	3.7	3.6	3.6	3.8	4.7	4.2	3.8	3.8	4.4
Map_HQ (4.9)	4.7	3.5	4.4	4.5	3.1	4.7	4.4	4.7	4.7	4.3	4.7	4.9	4.9	4.9	4.9

correlating with better MOS - which also applies to quality permutations where two media objects of the same quality are combined. However, in the case of quality permutations where media objects of different qualities are combined, the predicted MOS gravitates strongly towards the MOS of the higher quality object, rendering the MOS of the lower quality object being severely underrepresented.

B. OBM-specific model

To estimate QoE in OBM, all individual media objects need to be assessed. In other words, the QoE needs to be formulated as a collection of the individual QoE scores of separate media objects, where a single DASH model (such as P.1203) can be used to estimate the QoE of each object. We postulate this as:

$$QoE = \{P.1203(Object_1), \ldots P.1203(Object_n)\} \quad (1)$$

However, media objects can have varying levels of importance across different experiences as discussed in §IV-A. To combine individual QoE scores into a single metric, we need to find the balance between the individual media objects in terms of perceived quality by user. This might depend on the content type and will require subjective testing in the future. In the mean time we develop the following **four strategies** to quantify the impact of each media object on the overall QoE.

Average. This is a simple strategy in which we assume that all media objects are of equal importance in terms of the quality perceived by users. The overall QoE is calculated as:

$$QoE = Mean(P.1203(Object_1), \ldots P.1203(Object_n)) \quad (2)$$

Size. This strategy prioritises media objects according to their size in the final video element, where media objects occupying more of the screen have a correspondingly greater impact on QoE. The size of each object can be calculated by their average area of occupancy across the compositing canvas. The overall QoE is calculated as a weighted arithmetic mean of QoE scores of all media objects, with weights adjusted using their size.

$$QoE = \frac{\sum_{n=1}^{MaxObjects} P.1203(n) * Size(n)}{\sum_{n=1}^{MaxObjects} Size(n)} \quad (3)$$

Complexity. A strategy that leverages the Spatial Information (SI) of media objects, assuming that objects with higher SI will have greater impact on the perceived quality by users. SI is measured using the methodology outlined in ITU-T P.910, with overall QoE calculated as a weighted arithmetic mean of QoE scores of all media objects, with each score adjusted according to object's SI.

$$QoE = \frac{\sum_{n=1}^{MaxObjects} P.1203(n) * SI(n)}{\sum_{n=1}^{MaxObjects} SI(n)} \quad (4)$$

Motion. A strategy that prioritises media objects with higher amount of motion, measured as Temporal Information (TI) according to ITU-T P.910. The overall QoE is calculated as a weighted arithmetic mean of QoE scores of individual media objects, with scores adjusted according to measured TI.

$$QoE = \frac{\sum_{n=1}^{MaxObjects} P.1203(n) * TI(n)}{\sum_{n=1}^{MaxObjects} TI(n)} \quad (5)$$

VI. DISCUSSION OF RESULTS

To assess our model in the context of the weather forecast case study, we examine the P.1203 MOS of each quality permutation using the four strategies – see results in Table I.

Average strategy successfully represents the QoE of both media objects (map and presenter) as opposed to the baseline where the QoE estimate of each quality permutation was dominated by the higher quality object. *Size* strategy focuses on the resolution of media objects in the final video element, with larger objects being prioritised. In our example, the presenter and map media objects occupy 20% and 80%, respectively, of the final video's resolution. The resulting QoE estimates gravitate towards the MOS estimate of the map - the larger object. *Complexity* strategy prioritises objects with greater SI. In our example, the SI of the presenter and the map were 23.5 and 122.2, respectively. Similar to the second strategy, it resulted in the map object being better represented in the overall QoE, with differences in estimates between the two strategies within ± 0.1. In *Motion* strategy, media objects with greater TI have greater impact on the overall QoE estimation. In our example, the TI equalled to 11.6 and 6.3 for the presenter and the map media objects, respectively. This strategy resulted in the presenter object being better represented in the overall QoE, as opposed to the previous two strategies which resulted in the map object being prioritised.

All four strategies successfully improve over the baseline, resulting in a variety of approaches that led to both media objects being better represented. While the first strategy assumes that all of media objects are equally important in terms of QoE, the remaining three strategies prioritise objects by leveraging a variety of object characteristics: Size, SI and TI. However, these strategies are limited by the assumption that the relationship between the measured characteristic and the QoE importance is linear. The impact of these characteristics on the overall QoE needs to be investigated using subjective studies and quantified, which can vary across different types of programmes, to create a hybrid strategy.

References

[1] J. Carter, R. Ramdhany, M. Lomas, T. Pearce, J. Shephard, and M. Sparks, "Universal access for object-based media experiences," in *11th ACM Multimedia Systems Conference*, ser. MMSys '20. Association for Computing Machinery, 2020, p. 382–385.

[2] Z. Li, C. Bampis, J. Novak, A. Aaron, K. Swanson, A. Moorthy, and J. Cock, "VMAF: The journey continues," *Netflix Technology Blog*, vol. 25, 2018.

[3] N. Barman and M. G. Martini, "Qoe modeling for http adaptive video streaming–a survey and open challenges," *IEEE Access*, vol. 7, pp. 30 831–30 859, 2019.

[4] P. Georgopoulos, Y. Elkhatib, M. Broadbent, M. Mu, and N. Race, "Towards network-wide QoE fairness using openflow-assisted adaptive video streaming," in *SIGCOMM Workshop on Future Human-Centric Multimedia Networking (FhMN)*. ACM, Aug. 2013, pp. 15–20.

[5] C. Alberti, D. Renzi, C. Timmerer, C. Mueller, S. Lederer, S. Battista, and M. Mattavelli, "Automated qoe evaluation of dynamic adaptive streaming over http," in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2013, pp. 58–63.

[6] K. Yamagishi and T. Hayashi, "Parametric quality-estimation model for adaptive-bitrate-streaming services," *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1545–1557, 2017.

[7] M. N. Garcia, W. Robitza, and A. Raake, "On the accuracy of short-term quality models for long-term quality prediction," in *Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, pp. 1–6.

[8] J. Lievens, A. Munteanu, D. De Vleeschauwer, and W. Van Leekwijck, "Perceptual video quality assessment in http adaptive streaming," in *IEEE International Conference on Consumer Electronics (ICCE)*, 2015, pp. 72–73.

[9] K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of experience estimation for adaptive http/tcp video streaming using h.264/avc," in *IEEE Consumer Communications and Networking Conference (CCNC)*, 2012, pp. 127–131.

[10] H. T. T. Tran, T. Vu, N. P. Ngoc, and T. C. Thang, "A novel quality model for http adaptive streaming," in *IEEE Sixth International Conference on Communications and Electronics (ICCE)*, 2016, pp. 423–428.

[11] N. Staelens, J. De Meulenaere, M. Claeys, G. Van Wallendael, W. Van den Broeck, J. De Cock, R. Van de Walle, P. Demeester, and F. De Turck, "Subjective quality assessment of longer duration video sequences delivered over http adaptive streaming to tablet devices," *IEEE Transactions on Broadcasting*, vol. 60, no. 4, pp. 707–714, 2014.

[12] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. García, "Perceptual quality of http adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowdsourced subjective studies," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2141–2153, 2016.

[13] A. Asan, W. Robitza, I.-h. Mkwawa, L. Sun, E. Ifeachor, and A. Raake, "Impact of video resolution changes on qoe for adaptive video streaming," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 499–504.

[14] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the quality of experience of http video streaming," in *12th IFIP/IEEE International Symposium on Integrated Network Management (IM) and Workshops*, 2011, pp. 485–492.

[15] C. G. Bampis and A. C. Bovik, "Learning to predict streaming video qoe: Distortions, rebuffering and memory," *arXiv preprint arXiv:1703.00633*, 2017.

[16] D. Zegarra Rodríguez, R. Lopes Rosa, E. Costa Alfaia, J. Issy Abrahão, and G. Bressan, "Video quality metric for streaming service using dash standard," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 628–639, 2016.

[17] 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, 2017.

[18] D. Z. Rodriguez, J. Abrahao, D. C. Begazo, R. L. Rosa, and G. Bressan, "Quality metric to assess video streaming service over TCP considering temporal location of pauses," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 985–992, 2012.

[19] Y. Elkhatib, R. Killick, M. Mu, and N. Race, "Just browsing? Understanding user journeys in online TV," in *Conference on Multimedia*. ACM, Nov. 2014, pp. 965–968. [Online]. Available: http://doi.acm.org/10.1145/2647868.2654980

[20] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, and A. C. Bovik, "Modeling the time—varying subjective quality of http video streams with rate adaptations," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2206–2221, 2014.

[21] C. G. Bampis, Z. Li, and A. C. Bovik, "Continuous prediction of streaming video qoe using dynamic networks," *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 1083–1087, 2017.

[22] H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya, "Streaming video qoe modeling and prediction: A long short-term memory approach," *arXiv preprint arXiv:1807.07126*, 2018.

[23] ITU-T, "1203.3: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport-quality integration module," 2017.

[24] I. Recommendation, "910,"subjective video quality assessment methods for multimedia applications," recommendation itu-t p. 910," *ITU Telecom. Standardization Sector of ITU*, 1999.

[25] Apple, "HTTP Live Streaming (HLS) Authoring Specification for Apple Devices," 2022. [Online]. Available: https://developer.apple.com/documentation/http_live_streaming/http_live_streaming_hls_authoring_specification_for_apple_devices

[26] A. Raake, M.-N. Garcia, W. Robitza, P. List, S. Göring, and B. Feiten, "A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1," in *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, May 2017.

[27] W. Robitza, S. Göring, A. Raake, D. Lindegren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, M.-N. Garcia, K. Yamagishi, and S. Broom, "HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P.1203 – Open Databases and Software," in *9th ACM Multimedia Systems Conference*, 2018.

[28] S. Göring, W. Robitza *et al.*, "Itu-t rec. p. 1203 standalone implementation," 2018. [Online]. Available: https://github.com/itu-p1203/itu-p1203

[29] VideoLAN, "x264," 2022. [Online]. Available: https://www.videolan.org/developers/x264.html

[30] "FFmpeg," 2022. [Online]. Available: https://ffmpeg.org/