Speed Index: Relating the Industrial Standard for User Perceived Web Performance to Web QoE

Tobias Hoßfeld, Florian Metzger
Chair of Communication Networks
University of Würzburg
Würzburg, Germany
tobias.hossfeld@uni-wuerzburg.de

Dario Rossi
Networks for the Future
INFRES, Télécom ParisTech
Paris, France
dario.rossi@telecom-paristech.fr

***Abstract*—In 2012, Google introduced the Speed Index (SI) metric to quantify the speed of the Web page visual completeness for the actually displayed above-the-fold (ATF) portion of a Web page. In Web browsing a page might appear to the user to be already fully rendered, even though further content may still be retrieved, resulting in the Page Load Time (PLT). This happens due to the browser progressively rendering all objects, part of which can also be located below the browser window's current viewport. The SI metric (and variants) thereof have since established themselves as a de facto standard in Web page and browser testing.**

While SI is a step in the direction of including the user experience into Web metrics, the actual meaning of the metric and especially its relationship between Speed Index and Web QoE is however far from being clear. The contributions of this paper are thus to first develop an understanding of the SI based on a theoretical analysis and second, to analyze the interdependency between SI and MOS values from an existing public dataset. Specifically, our analysis is based on two well established models that map the user waiting time to a user ACR-rating of the QoE. The analysis show that ATF-based metrics are more appropriate than pure PLT as input to Web QoE models.

***Index Terms*—Web QoE Modeling, Page Load Time (PLT), Speed Index (SI), IQX model, WQL model**

I. Introduction

Reliably measuring Web users Quality of Experience (QoE) has proven to be equally challenging and useful. These challenges are clear considering the availability of numerous timing events from the browser[1], the implementation of which may very well differ between browsers. As such, simple events like the Page Load Time (PLT) are still broadly used to infer user QoE. The usefulness of measuring Web QoE is equally clear as a low QoE can directly translate into a loss of revenue[2], to the point that Google started using PLT to rank[3] search results [1].

However, it is far less clear how to distill an accurate estimation of subjective user QoE from the numerous browser signals. Figure 1 introduces a few of these events: the first byte of useful payload is received at the *Time to First Byte* (TTFB), whereas the document structure is fully loaded at the *Document Object Model* (DOM) time. These timescales are generally related to network latency and are too fast with respect to human perception. Rendering starts at the *Time to First Paint* (TTFP) and completes at the *Above-The-Fold* (ATF) time [2], i.e., when the current viewport is fully rendered — a timescale that is surely more relevant with respect to user perception. Finally, application-level *Page Load Time* (PLT) measures the precise time at which the page finishes loading, which includes content that is possibly not directly visible and that is thus less relevant for user perception. Others metrics apt at measuring user interaction have also been introduced in the past, such as the *Time to Interactive* (TTI), or the *Time to Click* (TTC). Notably, in the A/B tests performed in [3], the TTC informs us when users believe they have sufficient information to make a judgment on a perceived speed difference between the two Web pages being shown, and is thus a very good approximation of the ATF time — yet, TTC is only artificially available in controlled experiments.

Figure 1: Time-instant metrics when downloading and rendering a Web page. Above-the-fold (ATF) and page load time (PLT) indicate when the content shown in the visible part of the Web page ("above the fold") and when the page finishes loading, respectively. ATF is used to compute the speed index (SI), while PLT is traditionally used in different Web QoE models.

Ultimately, the experience of the user depends on the *whole process* of retrieving and displaying a Web page until the PLT has been reached [4]. But it especially depends on the browser rendering process up to the ATF time. To better capture the complete process, the Speed Index (SI)[4] metric was proposed by Google. It is defined as the integral of complementary visual progress (measured from histograms of pixel-level rendering, see Section III). Due to its computational costs, SI has been rarely used in practice, but it has raised the scientific community's interest. As a result, several SI variants have been introduced such as the Perceptual Speed Index (PSI), which uses Structural SIMilarity (SSIM) instead of pixel-level rendering, or ObjectIndex (OI) and ByteIndex (BI) [1], which approximate the rendering process by using simple object-level and byte-level completion ratios, respectively.

[1] https://www.w3.org/TR/2012/REC-navigation-timing-20121217/

[2] http://www.fastcompany.com/1825005/how-one-second-could-cost-amazon-16-billion-sales

[3] In particular, mobile page speed is used http://www.thesempost.com/google-mobile-first-index-page-speed-ranking/.

[4] https://sites.google.com/a/webpagetest.org/docs/using-webpagetest/metrics/speed-index

The SI (and its variants) represent a good step towards including the user experience into Web metrics (Sec. II). Yet, their actual meaning and especially their ability to predict user QoE are far from being fully elucidated. In this paper, we develop an understanding of the SI based on a theoretical analysis (Sec. III) and then use the SI as input to well established models (WQL and IQX) in order to map the user waiting time to the user QoE evaluation on an ACR scale (Sec. IV). Our results show that ATF-based metrics are more appropriate than pure PLT as input to Web QoE models, but there is still ample opportunity for future research (Sec. V).

II. Related work

QoE models for the Web are definitively not a new subject. Seminal work on the topic started with [5] and the recent adoption of new protocols, such as HTTP/2 and QUIC, refueled the interest on the topic. Special attention has been given to mobile Web browsing [6], [7], the impact of network bandwidth fluctuations and outages on Web QoE [8], [9], the impact of visual appeal [10] and usability [11] on QoE. Finally, a body of work studies waiting times for Web QoE models [1], [3], [12], [13]. In this work we are not considering session-based Web QoE [14], nor task-driven QoE [15], [16] or real-world distractions [17]. Instead the focus solely lies on the impact of Web page loading times, as, e.g., performed by Speed Index.

Broadly speaking, two classes of models can be used to estimate subjective user QoE from objective browser measurements. First, there are PLT-derived Web QoE models [4], [12], [18] that define QoE as a function $f(\cdot)$ of PLT and fit the design parameters to a dataset. And second, there are data-driven approaches [3], [6], [13] that use timings as features and employ machine-learning to learn the function from data. In this work, we focus on the former approach. In particular, two well known models that tie the PLT to QoE are built upon the IQX and the WQL hypotheses. IQX [19] is based on the assumption that, given a fixed stimulus t, the resulting change of QoE depends on the current level of QoE. Intuitively, the idea of IQX is that if the QoE is high, a small variation in the underlying QoS metric will strongly affect the QoE. This results in an exponential relationship between waiting time and QoE, yielding

$$QoE^{IQX}(t) = \alpha e^{-\beta t} + \gamma. \tag{1}$$

The WQL hypothesis [18] is instead based on the fundamental Weber-Fechner law from psychophysics and applied to waiting times. WQL assumes that the relationship between 'W'aiting time t and its 'Q'oE evaluation on a linear ACR scale is 'L'ogarithmic, and can be expressed as

$$QoE^{WQL} = a - b\ln(t). \tag{2}$$

The ITU-T G.1030 [18] model follows the WQL hypothesis and uses a logarithmic regression for a session[5] time t. ITU-T G.1030 specifies three models with parameters a,b that depend on the session time (short 6 s, medium 15 s, long 60 s sessions). Companies use this to measure and detect page load and to map PLT to MOS[6].

Independent of the specific mapping function, both the IQX and WQL models relate the user QoE to the user waiting time t, where t is approximated with the application-level PLT. Yet, it is well known that subjectively experienced time and objective physical time differ [20]. In other words, we expect a difference between *network-level*, *application-level* and *user-perceived* PLT. As outlined, in Web browsing a page might appear to the user to be already loaded although content is still being retrieved due to the progressive rendering of the browser and the fact that pages often stretch beyond the browser's viewport. This indicates that ATF might be a more appropriate stimulus than PLT as input to QoE models.

III. Definition and Understanding of Speed Index

A. Definition of Speed Index

For quantifying how fast a Web page is loaded over time, time-integral metrics are proposed which generally have the form of

$$M = \int_0^T (1 - R(t))dt = T - \int_0^T (R(t))dt. \tag{3}$$

$R(t)$ is the response of the Web page over time and indicates the completion ratio of the Web page, i.e. $0 \leq R(t) \leq 1$. As a consequence, since $R(t)$ is dimensionless, M refers to a time unit typically given in a scale of milliseconds $[\mathrm{ms}]$. The completion ratio can be computed in different ways. For the Speed Index $R(t)$ quantifies the visual progress by calculating the mean pixel histogram difference (MPHD) between the current Web page I_t at time t and the state of the page after ATF I_T, i.e. when the visible part of the page is completely rendered. Hence, the Speed Index with $T = \mathrm{ATF}$ is defined as

$$SI = \int_0^T (1 - R(t))dt, \quad R(t) = \mathrm{MPHD}(I_T, I_t). \tag{4}$$

For an image I, the histogram H_C for the colors $C \in \{R,G,B\}$ is a vector of n elements counting the number of pixels with the corresponding color code in, typically, $0, \ldots, 255$. The difference between the starting histograms (for the first video frame at time TTFP) and the ending histogram (last video frame at time ATF) is used as the baseline, namely

$$\varDelta H = \sum_{C \in \{R,G,B\}} \sum_{i=0}^{n-1} |H_{C,i}(\mathrm{ATF}) - H_{C,i}(\mathrm{TTFP})|. \tag{5}$$

Then the mean pixel histogram difference is the difference of the histogram for each frame in the video to the first histogram and compared to the baseline. This quantifies the *completeness* of each video frame as

$$R(t) = \frac{1}{\varDelta H} \sum_{C \in \{R,G,B\}} \sum_{i=0}^{n-1} |H_{C,i}(t) - H_{C,i}(\mathrm{TTFP})|, \tag{6}$$

[5] A session consists of three steps: (1) requesting, retrieving, and displaying of a search page; (2) typing and submitting a search term on this page; (3) retrieving and displaying of the results page.

[6] E.g. https://www.sandvine.com/downloads/general/sandvine-technology-showcases/web-browsing-qoe-score.pdf

with $R(t)=0$ for any $t<\mathrm{TTFP}$. Hence, SI is lower bounded by TTFB and upper bounded by T, $\mathrm{TTFP}\leq SI\leq T$.

The completion ratio $R(t)$ uses a full reference metric of the browser's output images at times t and T. Other full reference metrics lead to different time-integral metrics. The perceptual Speed Index (PSI) [3] uses SSIM [21] for $R(t)$ to better reflect user perceived quality (and provide resistance to visual jitter and shifts in the page layout during loading). PSI was introduced in 2016 to quantify how "most of the Web page's ATF content loads quickly without visually noticeable jitter" [3]. It is defined as

$$PSI=\int_0^T (1-R(t))dt, \quad R(t)=\mathrm{SSIM}(I_T,I_t). \tag{7}$$

The computation of the full reference metrics MPHD and SSIM are computationally expensive. Therefore, simpler metrics were proposed which use the ratio of bytes (*ByteIndex* or BI) or objects (*ObjectIndex*, OI) at time t compare to time T [1]. OI and BI use the relative amount of download objects, and bytes respectively, until the PLT. The downloaded bytes $B(t)$ over time are related to the overall volume V. BI can then be calculated by

$$BI=\int_0^T (1-R(t))dt, \quad R(t)=B(t)/V. \tag{8}$$

B. Understanding Speed Index

We start with a basic theoretical analysis of the Speed Index. If not mentioned otherwise, we consider the time instant $T=ATF$ when the visual progress is completed, i.e. $R(T)=1$. Therefore, it does not really matter what kind of SI variant we are looking at, since we consider only $R(t)$.

1) Smooth Visual Progress vs. Instantaneous Visual Completion: A (theoretical) linear visual progress is defined for

$$R(t)=t/T, \tag{9}$$

leading to a Speed Index of

$$SI=\int_0^T 1-R(t)dt=\left[t-\frac{1}{2}\frac{t^2}{T}\right]_0^T=\frac{T}{2}, \tag{10}$$

as depicted in Figure 2. For an instantaneous visual completion at time T, without any prior progress, the Speed Index is $SI=T$, i.e., the upper limit of the Speed Index (cf. also Figure 3). Hence, a smooth visual progress leads to half of the maximum SI. Since the visual progress can not start until the first byte arrived, the DOM is loaded and the first object (TTFP) is painted, we arrive at

$$\mathrm{TTFP}\leq SI\leq T=\mathrm{ATF}\leq\mathrm{PLT} \tag{11}$$

as an inequation for the Speed Index SI (or other derivates like PSI, BI, OI).

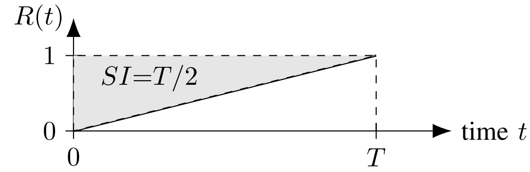

Figure 2: Continuous download and visual progress $R(t)=t$ until time T.

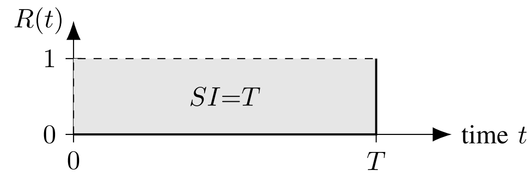

Figure 3: Visual progress is instantaneously completed at time T.

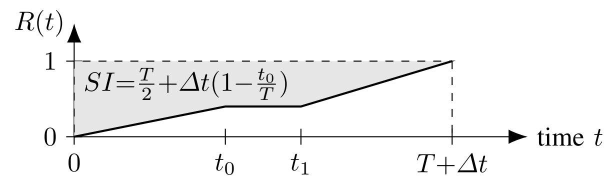

Figure 4: Smooth visual progress is interrupted at t_0 for time $\Delta t=t_1-t_0$. The progress is completed at time $T+\Delta t$.

2) Slowed Down vs. Interrupted Progress: But which kind of progression behavior leads to a better Speed Index? Let us consider a smooth visual progress which is interrupted at time t_0 for a duration of $\Delta t=t_1-t_0$, as shown in Figure 4. This increases the SI by the 'interruption rectangle' with length Δt and height $1-R(t_0)=1-t_0/T$. Thus, it is important to see that the time t_0 when the interruption starts is as important as the duration of the interruption due to the integral form of the SI. In contrast, ATF or PLT are only offset by Δt.

We compare the index SI_i of the interrupted progress with a slower, but smooth progress which also ends at time $T+\Delta t$ and has the same ATF. The slowed down progress yields $SI_s=\frac{1}{2}(T+\Delta t)$. This slowed down progress leads to a better QoE, i.e. a smaller SI, under the condition that

$$SI_s\leq SI_i \Leftrightarrow t_0\leq T/2. \tag{12}$$

Hence, if the interruption occurs early (i.e. before $T/2$), the interrupted progress is worse in terms of SI.

3) Time-discrete Progress: During practical observation the visual progress is measured as a time-discrete process. At n time instants t_i for $i=1,\cdots,n$, a certain progress R_i is observed. Thereby, an event at time t_i increases the visual progress by r_i. It is $R_1=r_1$, $R_n=1$, $T=t_n$, and $t_0=0$, i.e.

$$R_i=R_{i-1}+r_i=\textstyle\sum_{k=1}^{i} r_k. \tag{13}$$

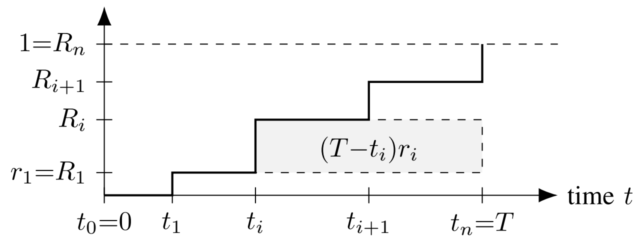

Figure 5: Time discrete visual progress. At time t_i, the progress is increased by r_i and the cumulative progress follows as $R_i = R_{i-1} + r_i = \sum_{k=1}^{i} r_k$.

For the SI during time-discrete progress then follows

$$SI = T - \sum_{i=1}^{n-1} (t_{i+1} - t_i) R_i = T - \sum_{i=1}^{n-1} (t_{i+1} - t_i) \sum_{k=1}^{i} r_k \quad (14)$$

$$= T - \sum_{i=1}^{n-1} (T - t_i) r_i. \quad (15)$$

4) Periodic Measurements and Slotted Time: Typical implementations of the SI periodically capture video frames to compute the visual completeness. For example, WebPagetest[7] currently captures at 10 frames per second. In general, we assume that there are n measurements within time $T = n\Delta t$. Hence, $\Delta t = T/n$ and $t_i = i\Delta t = iT/n$. Equation (15) leads to the SI which indicates that the first measurement at time t_1 is counted $(n-1)$ times.

$$SI = T - \sum_{i=1}^{n-1} (n-i) \Delta t \cdot r_i \quad (16)$$

Assuming a smooth, linear progress, we can set $r_i = 1/n$ and $R_i = R(t_i) = i/n$. Then, Eq. (14) leads to

$$SI = T - \sum_{i=1}^{n-1} \frac{T}{n} \frac{i}{n} = T - \frac{T}{n^2} \sum_{i=1}^{n-1} i = \frac{T(n+1)}{2n} \quad (17)$$

$$= \frac{T}{2} + \frac{T}{2n} = \frac{T}{2} + \frac{\Delta t}{2}. \quad (18)$$

Thus, the term $\frac{T}{2n} = \frac{1}{2}\Delta t$ indicates the absolute measurement bias due to the periodic discrete measurement points — instead of considering the continuous progress $R(t)$ yielding $SI = T/2$ as in Eq. (10). The relative bias is then $\beta = 1 + 1/n = 1 + \Delta t/T$. Figure 6 indicates the relative bias of the SI for different measurement periods Δt and smooth linear progress.

C. Summary

In practice, the Speed Index evaluates the visual progress of Web page contents. The measurement bias is not critical in practical implementations. When using ATF as end point for the computation of the SI, it can be interpreted as a proxy for the user perceived waiting time. This raises the question whether the WQL or IQX can be applied to the Speed Index.

[7] http://www.webpagetest.org/

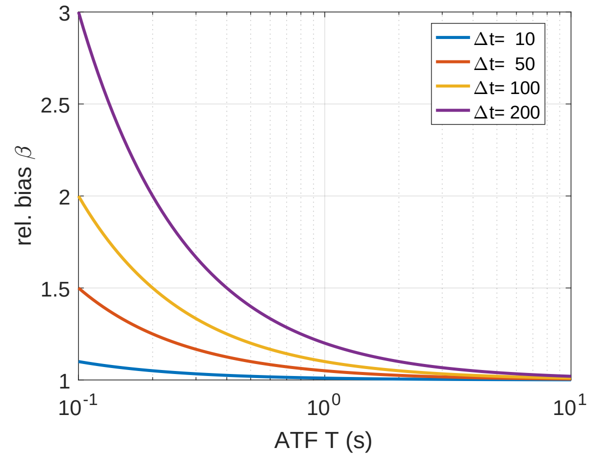

Figure 6: For the smooth linear progress, the absolute measurement bias $\Delta t/2$ of the SI is independent of the time T, but the relative bias is $1 + \Delta t/T$.

Table I: Characteristics of Web pages and the network characteristics in terms of PLT and ByteIndex in the subjective study.

	size (MB)	#objects	PLT (s)	BI plt (s)	BI atf (s)
mean	1.633	157	5.403	3.169	2.695
median	1.034	102	2.293	1.481	1.332
std	1.140	123	7.977	5.042	4.517
min	0.173	20	0.366	0.167	0.150
max	26.252	876	56.268	54.793	54.793
iqr	1.338	146	3.711	1.796	1.528

IV. Relationship between Speed Index and MOS

The authors of [1] provide a public database of Speed Index values and subjective ratings for many Web pages, which we use in the following to investigate the relation between MOS and SI. In particular, we suggest that the SI is a proxy for the user waiting time for the visible portion of a Web page to be rendered. Hence, we investigate the relationship between SI and MOS for WQL and IQX. Then, the Speed Index driven QoE model based on Eq.(1) and Eq.(2) is formulated as follows.

Web QoE (SI based QoE model SQ)**.**

$$SQ_{IQX} = \alpha e^{-\beta \cdot SI} + \gamma \quad (19)$$

$$SQ_{WQL} = -a \ln(SI) + b \quad (20)$$

A. Data Description

In [13], 8,687 Web browsing sessions were collected, wherein 241 volunteers rate their browsing experience with the Chrome browser. Twelve non-landing pages from the list of Alexa top 100 pages were used in the study. The size of the Web pages ranges from 0.17 MB up to 26.25 MB containing from 20 up to 876 objects, see Table I.

The users rated the Web QoE on a 5-point Absolute Category Rating (ACR) scale, with 5-Excellent, 4-Good, 3-Fair, 2-Poor, and 1-Bad. The dataset is published as open source database[8]. Beside the subjective ratings, it contains information about each page's page load time and several Speed Index variants,

[8] Web QoE dataset; https://newnet.telecom-paristech.fr/index.php/webqoe/

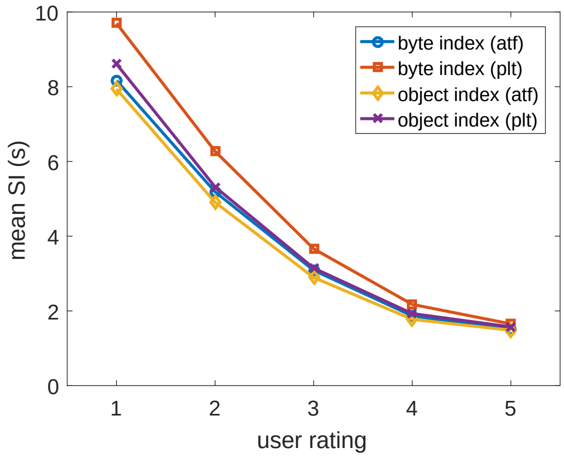

Figure 7: The average Speed Index is computed over Web pages and all users with a rating of $y \in \{1,2,3,4,5\}$.

based on both PLT and ATF. Details of the experimental setup are described in [1]. To be more precise, we use application-level PLT (indicated by 'onload' in the browser) as well as the Byte Index related to PLT as well as Byte Index related to ATF. The median PLT is 2.29 s, while the average PLT is about 5.40 s. The maximum observed PLT is 56 s. Accordingly, the BI values are in the same order, cf. Table I.

B. Numerical Results

The first observation is that no significant correlation between the individual user ratings and the different time instants as well as the Speed Index can be observed in the data. Therefore, we take a closer look at aggregated user ratings.

Figure 7 indicates the average Speed Index values for users rating a Web page with a value of y. We see a clear relationship between the user perception and the Speed Index. But there is no significant difference between BI (ATF) and OI (ATF and PLT). Nevertheless, the differences between BI PLT and ATF are significant. Figure 8 then shows the fitting between the BI ATF and MOS values according to the WQL and IQX hypothesis. This is performed by first grouping the BI values into bins of equal size, i.e. each bin contains the same number of subjective ratings; then the MOS is computed over the subjective ratings assigned into each bin. It can be seen that the simple WQL and IQX model lead to a very good fit. The goodness of fit values are provided in Table II. It is amazing how well both models captures the MOS when using the BI as input which provides a better fitting than using PLT instead. Figure 8 shows that there are not enough subjective ratings close to the edges. As a result, the MOS values do not drop below 2. The significant lack of samples in the public dataset does not allow to make conclusive considerations whether IQX or WQL is more appropriate.

Figure 9 compares the true MOS with the estimated MOS SQ when using the ByteIndex (ATF) as proxy in the WQL and IQX model for user perceived waiting time. This is compared to the application-level PLT. It can be clearly seen that ATF is required as proxy for user perceived waiting times. The

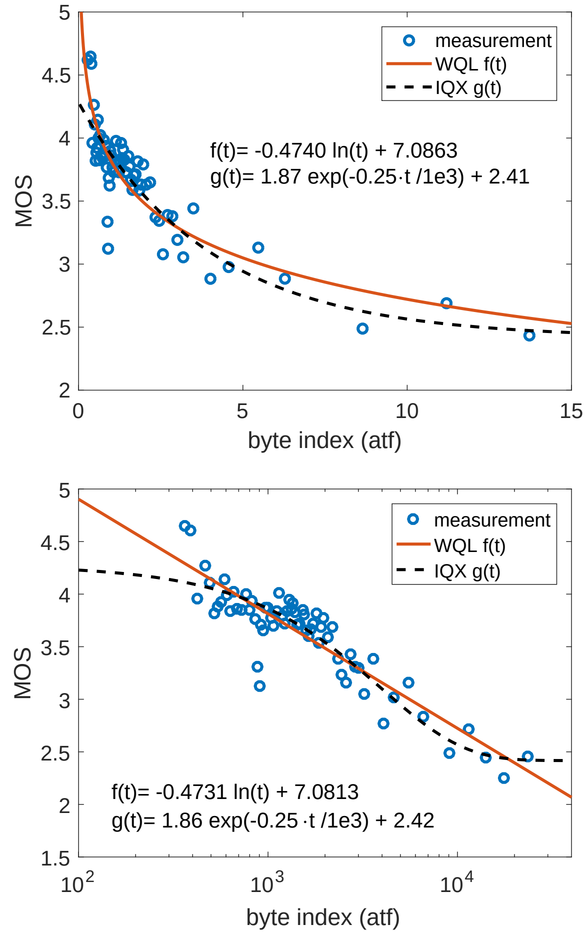

Figure 8: The WQL and IQX model leads to accurate results for Web page QoE when the byte index (ATF) is mapped to MOS. The logarithmic scale in the lower figure shows that there are however not sufficient user ratings at the edges.

Table II: Mean squared error (MSE), mean absolute error (MAE), correlation ρ, coefficient of determination ρ^2, maximum absolute error ($\max(AE)$) for the fittings with IQX and WQL

	WQL BI	IQX BI	WQL PLT	IQX PLT
MSE	0.053	0.070	0.168	0.132
MAE	0.168	0.160	0.175	0.183
ρ	0.919	0.891	0.741	0.792
ρ^2	0.844	0.794	0.549	0.628
$\max(AE)$	0.919	1.417	2.967	2.499

Speed Index nicely captures the visual progress and improves to quantify this user perceived waiting times. Then, again the WQL and IQX cannot be rejected and are promising models.

Previous subjective tests (e.g., [4]) employed less complex Web pages which yielded page load times that were almost identical to above-the-fold times. As a result, the WQL hypothesis was not rejected for page load times. The results here clearly indicate that ATF is more appropriate and that the byte index is a proper proxy for WQL. Figure 8 and Figure 9 show that the BI-WQL model leads to a very good estimation of the MOS values. Utilizing only page load times, on the other hand, does not lead to a good model, which is likely

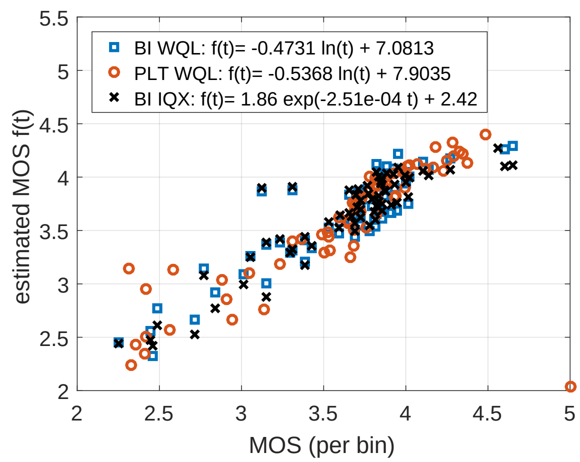

Figure 9: The estimated MOS from the byte index BI fits very well to the real MOS value. The model using page load times is not working so well.

caused by additional loading times for non-visible content ("below the fold") that do not affect users.

V. Conclusions and Future Work

This paper re-assesses the relevance of industrial standards, such as WQL-based ITU G.1030, for the evaluation of QoE Web users. In particular, due to the discrepancy between user-perceived and application-level page load time, we observe that novel metrics such at Above-The-Fold time are more appropriate than PLT. In particular, we can reinforce the notion that metrics in integral form, such as ByteIndex evaluated up until the ATF mark, yield a more accurate QoE evaluation. In other words, BI can be used as proxy for user perceived waiting times of above the fold Web page contents. Specifically, we show that WQL can be assumed under this requisite, logarithmically mapping BI to MOS values. Even IQX yields quantitatively similar accuracy results. The largest difference between IQX and WQL can be observed at very responsive pages (i.e., with BI in the range of $100\,\mathrm{ms}$ to $500\,\mathrm{ms}$), where there is a significant lack of samples in the public dataset, that do not allow to make conclusive considerations — although the IQX model yields more conservative estimations here.

While results in this paper are encouraging, they open a number of research questions. First, we observe that, while very useful, the dataset is still limited in terms of evenly covering MOS scale (i.e., few samples fall in the 4–5 range) and in the Web page diversity (only 25 pages are considered). This puts the generality of our findings into question. Enriching the dataset would therefore be a beneficial community-wide effort. Second, a larger dataset would allow to derive more precise models, i.e., by classifying Web sites into appropriate categories and determining models and parameters for each such class separately. Third, for the ByteIndex all bytes are born equal: integrating task-driven QoE by weighting relevant parts of a Web site to influence $R(t)$ is an obvious but interesting extension. The above directions are part of our future research agenda.

References

[1] E. Bocchi, L. De Cicco, and D. Rossi, "Measuring the quality of experience of web users," *ACM SIGCOMM Computer Communication Review*, vol. 46, no. 4, pp. 8–13, 2016.

[2] J. Brutlag, Z. Abrams, and P. Meenan, "Above the fold time: Measuring web page performance visually," in *Velocity: Web Performance and Operations Conference*, 2011.

[3] Q. Gao, P. Dey, and P. Ahammad, "Perceived performance of top retail webpages in the wild: Insights from large-scale crowdsourcing of above-the-fold qoe," in *Workshop on QoE-based Analysis and Management of Data Communication Networks*. ACM, 2017, pp. 13–18.

[4] S. Egger, P. Reichl, T. Hoßfeld, and R. Schatz, ""Time is bandwidth"? Narrowing the gap between subjective time perception and Quality of Experience," in *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1325–1330.

[5] J. Nielsen, "Response times: The 3 important limits," https://www.nngroup.com/articles/response-times-3-important-limits/.

[6] A. Balachandran, V. Aggarwal, E. Halepovic, J. Pang, S. Seshan, S. Venkataraman, and H. Yan, "Modeling web quality-of-experience on cellular networks," in *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014, pp. 213–224.

[7] S. Baraković and L. Skorin-Kapov, "Survey of research on quality of experience modelling for web browsing," *Quality and User Experience*, vol. 2, no. 1, p. 6, 2017.

[8] A. Sackl, S. Egger, and R. Schatz, "The influence of network quality fluctuations on web qoe," in *Quality of Multimedia Experience (QoMEX)*. IEEE, 2014, pp. 123–128.

[9] A. Sackl, P. Casas, R. Schatz, L. Janowski, and R. Irmer, "Quantifying the impact of network bandwidth fluctuations and outages on web qoe," in *Quality of Multimedia Experience (QoMEX)*. IEEE, 2015, pp. 1–6.

[10] M. Varela, T. Mäki, L. Skorin-Kapov, and T. Hoßfeld, "Towards an understanding of visual appeal in website design," in *Quality of Multimedia Experience (QoMEX)*. IEEE, 2013, pp. 70–75.

[11] M. Varela, L. Skorin-Kapov, T. Mäki, and T. Hoßfeld, "Qoe in the web: A dance of design and performance," in *Quality of Multimedia Experience (QoMEX)*. IEEE, 2015, pp. 1–7.

[12] S. Egger, T. Hoßfeld, R. Schatz, and M. Fiedler, "Waiting times in quality of experience for web based services," in *Quality of Multimedia Experience (QoMEX)*. IEEE, 2012, pp. 86–96.

[13] D. Da Hora, A. Asrese, V. Christophides, R. Teixeira, and D. Rossi, "Narrowing the gap between qos metrics and web qoe using above-the-fold metrics," in *PAM 2018-International Conference on Passive and Active Network Measurement*, 2018, pp. 1–13.

[14] T. Hoßfeld, S. Biedermann, R. Schatz, A. Platzer, S. Egger, and M. Fiedler, "The memory effect and its implications on web qoe modeling," in *Teletraffic Congress (ITC), 2011 23rd International*. IEEE, 2011, pp. 103–110.

[15] D. Strohmeier, S. Jumisko-Pyykkö, and A. Raake, "Toward task-dependent evaluation of web-qoe: Free exploration vs. "who ate what?"," in *Globecom Workshops (GC Wkshps), 2012 IEEE*. IEEE, 2012, pp. 1309–1313.

[16] D. Strohmeier, M. Mikkola, and A. Raake, "The importance of task completion times for modeling web-qoe of consecutive web page requests," in *Quality of Multimedia Experience (QoMEX)*. IEEE, 2013, pp. 38–39.

[17] D. Guse, S. Egger, A. Raake, and S. Möller, "Web-qoe under real-world distractions: Two test cases," in *Quality of Multimedia Experience (QoMEX)*. IEEE, 2014, pp. 220–225.

[18] E. Ibarrola, I. Taboada, R. Ortega *et al.*, "Web qoe evaluation in multi-agent networks: Validation of itu-t g. 1030," in *Autonomic and Autonomous Systems, 2009. ICAS'09. Fifth International Conference on*. IEEE, 2009, pp. 289–294.

[19] M. Fiedler, T. Hoßfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.

[20] S. Grondin, "Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions," *Attention, Perception, & Psychophysics*, vol. 72, no. 3, pp. 561–582, 2010.

[21] 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, pp. 600–612, 2004.