The Memory Effect and Its Implications on Web QoE Modeling

Tobias Hoßfeld, Sebastian Biedermann[1]
University of Würzburg,
Institute of Computer Science,
Chair of Communication Networks
D-97074 Würzburg, Germany
tobias.hossfeld@uni-wuerzburg.de

Raimund Schatz, Alexander Platzer, Sebastian Egger
Telecommunications Research Center Vienna - ftw
A-1220 Vienna, Austria
{schatz,platzer,egger}@ftw.at

Markus Fiedler
Blekinge Institute of Technology,
Communication and Computer Systems Research Lab
SE-371 79 Karlskrona, Sweden
markus.fiedler@bth.se

***Abstract*—Quality of Experience (QoE) has gained enormous attention during the recent years. So far, most of the existing QoE research has focused on audio and video streaming applications, although HTTP traffic carries the majority of traffic in the residential broadband Internet. However, existing QoE models for this domain do not consider temporal dynamics or historical experiences of the user's satisfaction while consuming a certain service. This psychological influence factor of past experience is referred to as the memory effect. The first contribution of this paper is the identification of the memory effect as a key influence factor for Web QoE modeling based on subjective user studies. As second contribution, three different QoE models are proposed which consider the implications of the memory effect and imply the required extensions of the basic models. The proposed Web QoE models are described with a) support vector machines, b) iterative exponential regressions, and c) two-dimensional hidden Markov models.**

I. Introduction

User satisfaction with application and service performance in communication networks has attracted increased attention during the recent years. Parts of this growth of interest in *Quality of Experience* (QoE) issues can be explained by increased competition amongst providers and operators, and by the risk that users churn as they become dissatisfied [1]. Users do not like to wait unnecessarily; they might consider a slow interactive service as worthless [2], but find it worthwhile to tell other users about bad experiences [1]. As the risk for anger, gossip and churn grows with the time a user is dissatisfied, it becomes important to develop the understanding of the *dynamics of user perception* in form of a *QoE model*.

After years of dominance of peer-to-peer traffic, interactive HTTP traffic again took the pole position in residential broadband Internet traffic with a share of more than 50% as found by a recent study [3]. One might well imagine millions of people sitting in front of their browsers and waiting for their YouTube videos and social networking tasks to be executed. Many users face volatile network conditions due to signal-to-noise ratio problems on wireless or DSL links or temporary over-utilization of shared network resources. So far, the reactions of users to waiting times on the web are well-researched and well-described in QoE models [4]–[6]. Still, we do not know about transitions of user experience as waiting times change. Classical QoE models lack such dynamical components.

Most work in the context of QoE modeling is done for multimedia applications like voice or video services. Here, the majority of QoE models focus on the current stimuli, i.e. the actual service environment and conditions, and do not consider the temporal dynamics or historical experiences of the users' satisfaction while consuming a certain service. This psychological influence factor of past experience is referred to as the *memory effect*. Therefore, the goal of this paper is to derive a specific QoE model for web traffic considering temporal dynamics and the memory effect and henceforth called *Web QoE model*.

In order to derive a Web QoE model, three steps are proposed. Firstly, subjective user studies are conducted in which influence factors on the user perceived quality are varied. The network conditions are emulated such that users experience a page load time (PLT) that varies slightly around a predefined value. Secondly, the key influence factors (KIFs) are identified by means of statistical analysis and machine learning methods. As a result of this paper, the PLT as well as the memory effect are recognized as KIFs. Finally, three different QoE models are proposed which extend basic models due to the memory effect: a) support vector machines, b) iterative exponential regressions, and c) two-dimensional hidden Markov models.

The remainder of this paper is structured as follows. Section II gives an overview of related work. We focus on literature quantifying QoE for web traffic, as well as on interdisciplinary research studies from human-computer interaction about remembered user experience. The design of the subjective user studies and the measurement setup are explained in Section III, before we present the statistical analysis of selected KIFs in Section IV. The derived Web QoE models are introduced in Section V. Finally, Section VI concludes this paper with the main finding and gives an outlook on our next steps for Web QoE modeling.

The work has been supported by COST Action IC0703, the project G-Lab, funded by the German Ministry of Educations and Research (Förderkennzeichen 01 BK 0800, G-Lab), and by the European FP7 ICT No 216366 "Euro-NF" through the Specific Joint Research Project "QoEWeb". The authors alone are responsible for the content of the paper.

[1]Sebastian Biedermann is now with Technische Universität Darmstadt, Germany (biedermann@seceng.informatik.tudarmstadt.de).

II. Related Work

This section discusses existing work in the domain of QoE Modeling for the web as well as related work on QoE modeling for temporal phenomena such as the memory effect.

A. Web QoE Modeling

In this paper we define Web QoE as *Quality of Experience of interactive services that are based on the HTTP protocol and accessed via a browser*. The most prominent examples for applications of this category are surfing the web and downloading files, e.g. mp3 songs. In the context of web browsing it has been widely recognized that in contrast to the domains of audio and video quality, where psycho-acoustic and psycho-visual phenomena are dominant, end-user *waiting time* is the key determinant of QoE: the longer users have to wait for content to arrive or transactions to complete, the more dissatisfied they become with the service [7], [8]. Accordingly, user frustration significantly increases when waiting times without feedback exceed 8–10 seconds [2], [9]. Consequently, a number of studies have been conducted with the goal to quantify the relationships between Web QoE, application-level metrics (such as response and download times) and the QoS of the underlying network.

On the application level, ITU-T Recommendation G.1030 [4] discusses the results of user experiments that relate QoE ratings in terms of MOS (Mean Opinion Score, cf. [10]) of test subjects to different session times (i.e., aggregated response times) in the context of a simple web search task. The resulting model approximates Web QoE as logarithm of the normalised session time. G.1030 was validated by [11] who conducted extended versions of aforementioned experiments that address contemporary network speeds and changing user expectations. A logarithmic relationship between download time and QoE was also identified in [5] and explained by fundamental laws of psychophysics. Other studies have quantified the links between network QoS and Web QoE using similar methods. For example in [6] and [12], results from network-level measurements in the context of web user experiments were correlated with subjective quality ratings in order to model the impact of QoS parameters such as packet loss and delivery bandwidth on QoE.

What aforementioned Web QoE studies have in common is that they directly relate current system conditions (such as load time, bandwidth, packet loss) to MOS scores using mapping functions of the form $MOS = a + b \cdot f(QoS)$. Consequently, the resulting models all exhibit the same shortcoming: they are stateless since they only take *current* system and environment conditions into account without considering the impact of past conditions and experiences on the subjects' quality judgement. However, as discussed in the following subsection, such influences in the form of memory effects have been shown to exert significant influence on end-user quality perception and also have been already successfully integrated in QoE models for audio and video services.

B. QoE Modeling Considering Experience over Time

Due to the increased importance of packet switched networks for media delivery and the resulting temporal fluctuations of media transmission quality, the time-dynamics of experience has become a central topic in audio-visual quality research during the last decade. As regards multimedia QoE, typical temporal fluctuations of media quality take place within time spans (between 15sec up to several minutes [13], [14]) that are primarily covered by *short-term memory* (STM) and *working memory*, which is based on the interplay between STM and controlled attention, respectively [15], [16]. Research on the STM [17] has revealed the influence of primacy and recency effects of stimulus presentation on the human ability to memorize certain stimuli. These memory effects have also been studied in the context of audio [13], [18] and video QoE [19], resulting in their quantification as time constants of changes in perceived QoE caused by media quality improvement or degradation. These results have been successfully integrated in objective QoE prediction models such as [20], [21] where peak impairments and aforementioned time constants were used to model exponential decay or rise of the user perceived quality as reaction to media quality drops or rises. In the field of audio quality, a filter for recency modeling has been approved as standard by ITU-R [22].

As mentioned previously, current Web QoE modeling approaches only consider temporality in form of waiting times. They do not take into account memory effects such as recency effects. Similar to the approaches in audio and video QoE, it is necessary to conduct studies to identify a) the impact of the memory effect and b) determine memory time constants suitable of Web QoE tasks. These studies are necessary since the nature of the experience of interactive web applications differs strongly from pure audio and video quality experiences in media consumption. Based on these dedicated studies' results, memory effects could then be incorporated in Web QoE models in order to enhance their prediction accuracy.

III. Design of User Tests and Measurement Setup

In this section, we introduce the general categories of influence factors on Web QoE and how they relate to the design of our subjective user tests. Furthermore, we discuss the measurement setup and instrumentation of the three user experiments we conducted.

A. QoE Influence Factors and Design of User Tests

Since existing Web QoE studies did not consider temporal influence factors such as the memory effect, dedicated subjective user experiments had to be designed and conducted in order to identify the relevant influence factors, quantify their impact on the QoE and develop appropriate models. In general, a variety QoE influence factors exists on different levels:

a) technical level, e.g. network delivery bandwidth, page load time, packet loss, browser type,
b) psychological level, e.g. expectations regarding quality levels, type of user, or memory effects,

c) content level, e.g. type of website, design or implementation of websites,
d) context level, e.g. physical location or social context.

In our experiments, we focused on the technical and the psychological level since the primary goal was to investigate whether the experiencing subject's internal state (i.e., the psychological level) can be as influential as the current technical conditions and thus also needs to be considered in QoE modeling.

On the technical level, we subsumed the influences of network, application, etc. under a single variable: the *page load time* (PLT), defined as the time from issuing the request until the page is fully loaded and displayed. For simple web pages this aggregation is valid, since the most relevant technical influence factors on the transmission of web traffic (such as packet delay, packet loss, throughput) directly affect the transfer times of HTTP objects, which in turn determine the overall PLT. This way, the test parameter space was significantly reduced as necessitated by the generally limited number of test conditions that can be covered in a user test session. In order to further reduce the number of test cases required, we did not vary influence factors on content level or compare between different test designs. For the content, we used a simple photo webpage displaying a single image in order to avoid any content specific influences on user quality perception and rating behavior.

On the psychological level, we tested for the memory effect on behalf of the following test design: each test user encountered a series of web pages. After the download of each web page, the user was prompted for his or her opinion about the overall QoE on a given rating scale. Thus, the participant rated the QoE multiple times during a web session on the same server. In order to investigate the memory effect, all participants experience the same defined sequences of PLTs in the online tests. Such *identical sequences* are in contrast to typical ITU-T recommendations, which suggest the usage of *randomized sequences* in order to cancel out unwanted interferences between subsequent test conditions (cf., [23]–[25]). However, since we exactly wanted to investigate such interferences, i.e. the memory effect, identical PLT sequences were required.

For quantifying the QoE we used Absolute Category Rating (ACR), a common test method used in quality tests and standardized in ITU-T Recommendation P.910 [24]. According to this method, a single test condition (in our case the download of a website with a preset PLT) was presented to the test user who then issue a quality rating on an ACR scale. To this end, we used a 5-point MOS scale [10] as described in Table I.

TABLE I
ITU-T 5-POINT SCALE FOR ABSOLUTE CATEGORY RATING [10]

Grading Value	Estimated Quality	Perceived Impairment
5	excellent	imperceptible
4	good	perceptible but not annoying
3	fair	slightly annoying
2	poor	annoying
1	bad	very annoying

B. User Test Setup and Implementation

We implemented two different measurement setups: a local testbed as well as an online test. Both measurement setups allowed to preset a constant page load time, the independent variable, for each web site. Furthermore, the user ratings were stored in a database as well as the traffic traces in TCP dump files. The general advantage of a local testbed is that test users can be observed during the test run, in order to e.g. manually identify the behavioral category the participant belongs to. However, this lab-based approach is very time consuming and resource intensive, thus only a relatively small number of users can be covered. In contrast, online tests allow for a much larger number of participants. However, no direct observation of users and their reactions is possible.

During a lab test session using the local testbed, a single user was sitting in front of a personal computer and sequentially downloaded web pages. Each web page contained a randomly chosen image downloaded within the predefined PLT. In the local testbed, the PLT was adjusted by using the NistNET traffic shaper [26] which added a defined amount of delay to the passing IP packets requested. For the online test, the participant interacted with a Java applet that already contained the contents of the websites. The applet simulated the download of various web pages with predefined PLTs. The web page also contained MOS rating buttons from 1 to 5 according to Table I which were used by the test user to give his current personal satisfaction rating during the browsing session. In particular, subjects were asked to answer the question *"Are you satisfied with this download speed?"*. Finally, the user ratings issued were stored in a central MySql database.

C. Conducted Experiments

In the remainder of this paper, we will analyze the results from three different experiments, referred to as *local test* being conducted in the local testbed and two online tests referred to as *online test #1* and *online test #2*. The three experiments differ in the sequence of instrumented PLTs for the web pages as depicted in Figure 1. It has to be noted that each PLT was applied to at least four web pages in a row in order to allow for detection of the influence of the memory effect and when the effect diminishes. Since we used a traffic shaper to determine the PLT, we captured the network traffic in order to measure the real download time. In the online test, the preset PLTs did not deviate from the real download time, since the applet was executed on the test participant's local machine.

The local test comprised of 93 web pages and was conducted by 29 users. During the online tests, a user viewed 40 web pages. There were 72 and 26 users completing the online test #1 and #2, respectively. In the local test, we used PLTs up to 8 s, while in the online tests the maximum PLT was only 1.2 s in order not to scare the online users away due to long waiting times and an accordingly frustrating user experience. Details of the three tests can be found in Table II.

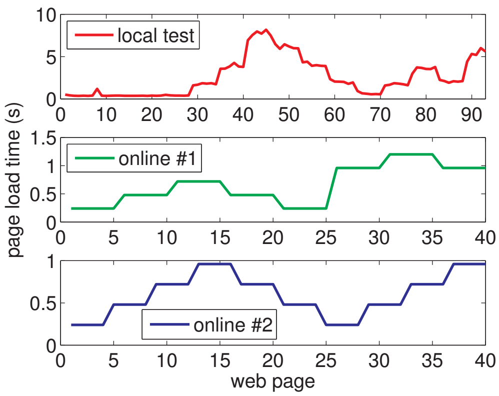

Fig. 1. Local and online user tests in which the page load times of the consecutively downloaded web pages were varied.

TABLE II
TEST DESIGN AND QOS SETTINGS IN TERMS OF PAGE LOAD TIME FOR THE DIFFERENT USER STUDIES

experiment	#users	#pages	min	max	mean	CoV
local	29	93	0.35 s	8.00 s	2.57 s	0.88
online #1	72	40	0.24 s	1.20 s	0.66 s	0.51
online #2	26	40	0.24 s	0.96 s	0.60 s	0.42

IV. STATISTICAL ANALYSIS OF SELECTED INFLUENCE FACTORS

In this section, we statistically analyze the impact of selected influence factors on Web QoE in order to identify the KIFs for deriving appropriate web QoE models. In particular, we consider the page load time and the memory effect as discussed in the design of user tests in the previous section. It has to be noted that due to space limitations we do not present the analysis of additional KIFs like the type of user or the content of a website.

A. Page Load Time (PLT)

As already discussed in the previous two sections, the PLT well aggregates the influences of network transmission on Web QoE, directly relating to end-user waiting time. Figure 2 illustrates this influence by showing the results of online test #1. The sequence of the 40 web pages downloaded by the test user is plotted along the x-Axis. On the left y-axis, the share of users rating the QoE in each category OS_1 (bad) to OS_5 (excellent) of the 5-point ACR scale is illustrated as stacked bar plot. Thus, the bars represent the relative distribution of participants' opinion scores for each page of the test sequence.

In addition, Figure 2 displays the MOS per web page, which is the average over all user ratings for this web page, as well as the instrumented PLT in seconds (as scaled on the right y-axis). Overlaying both curves reveals the inverse relationship between the PLT and Web QoE: the higher the PLT set for a webpage, the lower the resulting MOS score. For online test #1, Pearson's correlation coefficient between PLT and MOS is $-0{,}957$, thus PLT and QoE are indeed tightly coupled.

As next step, we model the impact of the PLT on the MOS. In [5], the logarithmic nature of QoE for a given QoS

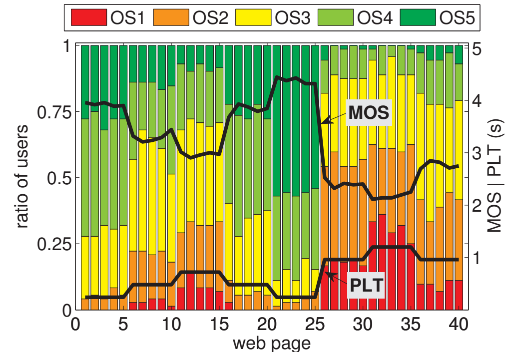

Fig. 2. Cumulative distribution function of user ratings for online test #1.

parameter like the PLT is discussed because of the underlying Weber-Fechner Law (WFL) from psychophysics. The WFL is used to describe the relation between the magnitude of a physical stimulus and its perceived intensity with the human sensory system. It assumes the differential perception dP to be directly proportional to the differential change $\frac{dS}{S}$, i.e. $dP \sim \frac{dS}{S}$, which can be solved as logarithmic relationship

$$P = k \ln \frac{S}{S_0} \tag{1}$$

with S_0 as threshold of the stimulus below which it is not perceived at all.

In the context of Web QoE, the stimulus S is the PLT and the perceived intensity P is the QoE. The applicability of the WFL for the user experiments conducted in this work is validated in Figure 3 which shows the MOS (averaged over all users) for each web page in dependence of the PLT for the three experiments. The x-axis is scaled logarithmically, such that each instance of WFL-based model appears as solid straight line in Figure 3. The parameters k and S_0 of each model instance are determined using nonlinear regression by minimizing the least-squared errors between the model function in Equation (1) and the MOS of the user ratings depending on the PLT. It can be seen that there is a good match between the logarithmic approximation and the measurement results, with coefficients of determination D being 0.85 and higher. Particularly the results for the online tests lead to very similar model instances, i.e. overlapping lines, since the instrumented PLTs are in the same order of magnitude.

The deviations of the MOS test data from our logarithmic models in Figure 3 can be explained by the fact that, the Weber-Fechner law does not consider temporal dynamics, but only the current stimulus. Thus, the QoE experienced in the past is not taken into account by our WFL-based models, similar to the other models currently used in the Web QoE domain.

From these observations we conclude that additional, time-related influences such as the memory effect might impact on the QoE as strongly as technical influences (as represented here by the PLT). Consequently, we will further investigate the memory effect in the following section.

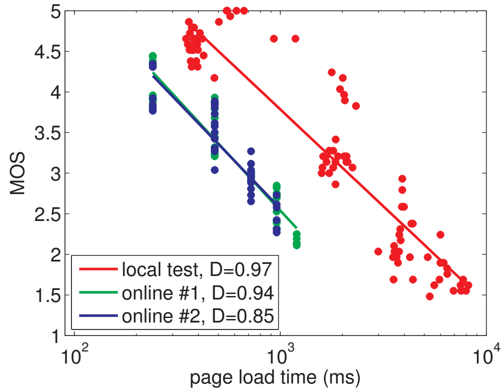

Fig. 3. Scatter plot of the MOS for each web page for the three tests. The WFL-based model instances are represented as solid line, while the MOS (averaged over all users) computed for each page are depicted as dots.

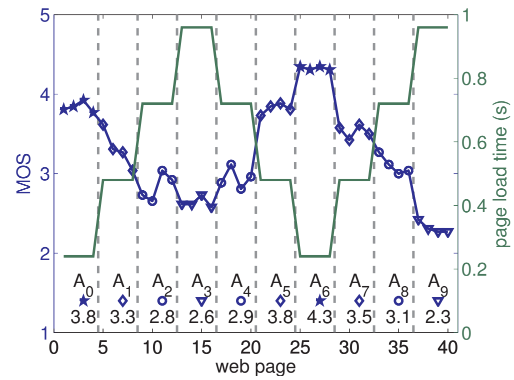

Fig. 4. Memory effect for average user ratings for online test #2.

B. Memory Effect

As potential key influence factor on Web QoE, the memory effect itself is going to be analyzed. As we have already seen in Section IV-A the current PLT alone is not sufficient for predicting the resulting MOS, which suggests the presence of the memory effect. However, the question remains whether the memory effect is relevant and significant enough to justify its integration in Web QoE modeling.

The answer to this question is provided by analyzing the time series of average user ratings. Figure 4 shows the MOS over all users for the 40 web pages of the test sequence of online test #2. In addition, the PLT is depicted on the right y-axis. In the figure, pages downloaded with identical PLTs use the same marker. The stars and the diamonds represent the MOSs for web pages with a PLT of 240 ms and 480 ms respectively. Since the PLT value was always held constant for groups of four web pages in a row in online test #2 to allow for investigating the influence of the memory effect, we have ten different sections A_i of four consecutively downloaded pages. The average MOS μ_i for each group of pages with identical PLT is additionally printed below the marker symbol of the respective section.

The memory effect manifests itself in two ways: transient effects and different average MOS levels for different page groups (sections) featuring the same QoS (PLT). Firstly, we consider transient effects by analyzing the evolution of the MOS over the four consecutive pages within a single section A_i. If there is a decrease of the service quality, i.e. $\mathrm{PLT}(A_i) > \mathrm{PLT}(A_{i-1})$, then, the MOS strongly decays and this decay continues for the other web pages in this section A_i. We assume here an exponential decay similar to [20], which can be clearly identified for some sections, e.g. A_8 or A_9. This observation is taken into account later for the iterative exponential regression QoE model in Section V-B.

Secondly, we compare two different sections A_i and A_j with the same PLT. For example, the PLT in section A_1 and A_5 is set to 480 ms. However, the average μ_i over the MOS values of the four consecutively downloaded web pages differs significantly in section A_1 and A_5 with $\mu_1 = 3.3$ and $\mu_5 = 3.8$ respectively. The difference of the ratings lies in the 95 % confidence interval $[0.26; 0.76]$. The reason for this difference is the memory effect, as the perception of the current QoS level is strongly biased by the preceding quality levels experienced. The PLT in A_0 is 240 ms and then the users experience a worse quality in A_1 with a PLT of 480 ms. In A_4, the PLT is 720 ms and then quality is increased in A_5 with a PLT of 480 ms. Thus, the users assume that quality has actually improved and thus give a better rating in A_5 than for the same QoS in A_1.

It has to be noted that the impact of the memory effect differs for the various sections with same PLTs. E.g., in section A_2 and A_4 the average of the MOS ratings per section μ_2 and μ_4 differ only about 0.11. We still have to show statistically that the memory effect is relevant and a true KIF. This will be done in Section V-A where support vector machines are used as QoE model and as a side effect the memory effect is identified as KIF in a statistically robust way.

V. Quality of Experience Models

Three different Web QoE models are presented in the following that are A. Support Vector Machine (SVM), B. Iterative Exponential Regression Model (IERMo), and C. Hidden Memory Markov Model (HMMM). The SVM is used for classification into QoE categories based on the identified KIFs. It predicts into which QoE category a new example, i.e. a set of KIFs, falls. The IERMo allows to directly calculate a MOS for a given sequence of PLTs. The HMMM describes the evolution of hidden system states and the observed emission from such a hidden state is the QoE grading value of an individual user. We highlight the implications of the memory effect on the QoE models, i.e., how to include the memory effect in the models and discuss the structural changes of the basic models.

A. Support Vector Machines

Support Vector Machines are one choice to make a model for classification with identified KIFs. One advantage is that every variable gets a weight from the model indicating as weight of importance, if there are not differently dependent, correlating and independent variables mixed. The disadvantage is that SVMs are acting on two-class-problems. For this we

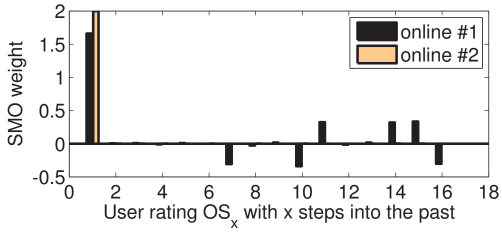

Fig. 5. Weights of the SMO models for the user online tests #1 and #2 with user ratings $\mathcal{OS}_x$ from past, respectively.

take the categories 1 to 3 of the ACR scale to *class −1* and the categories 4 to 5 to *class 1*. For practical reasons we choose the implementation of SMO (Sequential Minimal Optimization [27]) in WEKA [28] for analysis.

As input variables for the SMO, we use the current page load time PLT and the past values of the user rating $\mathcal{OS}$. Thus, the previous user rating is considered by $\mathcal{OS}_1$, the user rating one step more in the past $\mathcal{OS}_2$ and so on. With m users conducting a test run consiting of n websites, we arrive at $m{\cdot}n$ input samples for the SMO.

As a result, we got the $\mathcal{OS}_x$ weigths of the SMO models for the user online tests #1 and #2 in Figure 5. We see that $\mathcal{OS}_1$ has a strong weight, i.e., the last user perceived quality strongly influences the current user ratings. In particular, the memory effect is as important as the current PLT, as the weights are in the same order. Thus, the memory effect exists and is relevant for Web QoE modeling. We see that further steps into the past, i.e. $\mathcal{OS}_j$ for $j > 1$ are not relevant and we only have to take into account the last user rating.

The class $C \in \{-1, 1\}$ of the user rating can be predicted using these weights of all input variables V. Class -1 means that the user is dissatisfied corresponding to fair, poor, or bad quality, while class 1 means a satisfied user corresponding to good or excellent quality: $C = \text{sign}\left(\sum_{\{\text{variables } V\}} w_V \cdot V\right)$, where w_V is the corresponding normalized weight of the SVM.

The evaluation of the SVM model is done using 10-fold cross validation of the measurement data for the two online tests with the $\mathcal{OS}_x$ past series of user ratings. For online test #1, the true prediction rate is 86.8 % at a lower limit of 62.2 %. Thus, the SVM model correctly predicts 86.8 % of the user ratings, which significantly improves the lower limit as retrieved by selecting always the more frequent class C. For online test #2, we observe a true prediction rate of 82.3 % at a lower limit of 50 %. The cross validation with the respective other test gives similar values. Thus, the results of the SVM models relying on $\mathcal{OS}_x$ are consistent for both tests.

The main finding of the SVM model is that the memory effects exists, but only the user perceived quality of the last downloaded web page has to be taken into account.

B. Iterative Exponential Regression QoE Model

The *Iterative Exponential Regression Model (IERMo)* receives the PLT_i for a sequence of i web pages as input parameter and estimates the mean opinion score MOS_i for a specific user group. As a result of the SVM models, we consider only the short-term memory effect, i.e. for the computation of MOS_i we only take into account the previous MOS_{i-1}. Similar to [20] where the evolution of VoIP quality is estimated by means of exponential regressions, the fundamental equation of the IERMo is as follows.

$$\text{MOS}_i = f(\text{PLT}_i) - \omega e^{-j} \cdot (\text{MOS}_{i-1} - f(\text{PLT}_i)) \qquad (2)$$

Thus, the current MOS_i depends on the previous experience MOS_{i-1} and the actual PLT_i which is mapped according to the Weber-Fechner law in Equation (1) to the corresponding perceived intensity $\text{MOS}_i = f(\text{PLT}_i)$. It has to be noted that f is a time-independent function. The evolution over time is taken into account by exponential regressions, such that the memory effect decays exponentially over consecutive web pages with similar PLTs. In particular, for the j-th web page in a row with similar PLTs, i.e. $|\text{PLT}_{i-1} - \text{PLT}_i| < \epsilon$, the quality difference between the previous MOS_{i-1} and the currently perceived intensity $f(\text{PLT}_i)$ is only considered with ωe^{-j}. Otherwise, the user perceive a new test condition, i.e. $|\text{PLT}_{i-1} - \text{PLT}_i| \geq \epsilon$, and the previous quality is taken into account with $j = 1$.

The parameter ω is retrieved by minimizing the least-squared errors between the measured data of online test #1 and the resulting model values MOS_i. In particular, we used the mean absolute relative error Θ which was found to follow

$$\Theta(\omega) = 0.019\omega^2 - 0.0095\omega + 0.052\,, \qquad (3)$$

and which achieves the minimum for $\omega = 0.254$. This model can be further enhanced by considering the user types individually, since the memory effect is distinctive according to the type of user. For different user types, we have different functions f and weights ω. For example, insensitive users do not show a memory effect, i.e. $\omega = 0$, since the users are more or less always satisfied, while for hectic users the memory effect is strongly observed and accordingly reflected by the weight ω.

An evaluation of the IERMo is illustrated in Figure 6. As training data, we used the online test #1 to obtain the parameter ω. Applying iteratively Equation 2 leads to the following mean absolute relative errors of $\delta = 10.00\,\%$ for the local test and $\delta = 6.18\,\%$ for the online test #2, respectively. Applying the IERMo to the training set leads to $\delta = 5.94\,\%$ for the online test #1.

Summarizing, the IERMo is a simple, but efficient algorithm for Web QoE modeling which takes into the memory effect with an exponential decay for different types of users.

C. Hidden Memory Markov Model

Another approach to model the time-dynamics of QoE caused by the memory effect is a two-dimensional Hidden Markov Model. The hidden states describe the internal system state while the emission describe the observed user ratings. As a result of the SVM model, it is sufficient to consider the previous web page in order to take into account the

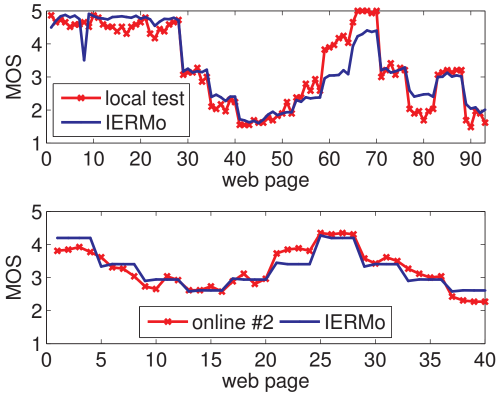

Fig. 6. Iterative exponential regression QoE model for local test and online test #1 with mean absolute relative error $\delta_1 = 5.18\,\%$ and $\delta_2 = 6.66\,\%$.

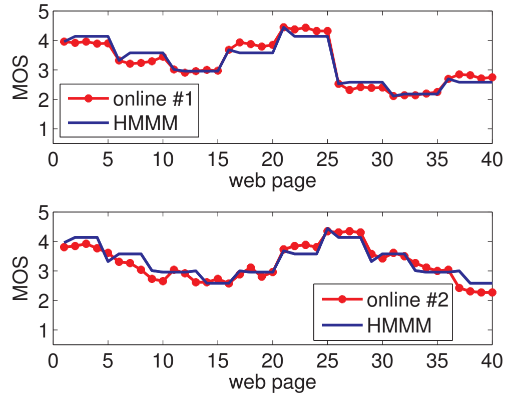

Fig. 7. Hidden Memory Markov QoE model for online test #1 and #2 with mean absolute relative error $\delta_1 = 4.49\,\%$ and $\delta_2 = 6.24\,\%$ respectively.

memory effect properly. Therefore, the hidden states do not include only the current PLT, but also a second state to capture the previous download time. This way, we obtain a Memory Markov model similar to [29] to "remember" the past system states. The emission from a hidden state reflects the individual user rating on a certain rating scale like ACR. Thus, the model is referred to as Hidden Memory Markov Model (HMMM).

A sequence of N web pages with corresponding PLTs $\{x_1, x_2, \ldots, x_N\}$ is now extended to a series of pairs $\{(x_i, x_{i-1})\}$ for $i = 2, \ldots, N$. Hence, the series of pairs consists of one sample less than the original time series $\{x_i\}$. Each pair of the extended series $\{(x_i, x_{i-1})\}$ is related to a hidden system state (H_i, H_{i-1}) where $H_i \in \{1, \ldots, M\}$ denotes the current PLT class and $H_{i-1} \in \{1, \ldots, M\}$ the previous PLT class. The classes are obtained by discretizing both the x_i and x_{i-1} as follows.

$$H_i = \begin{cases} 1 & \text{if } x_i = \min_i x_i \\ \left\lceil \frac{x_i - \min_i x_i}{\max_i x_i - \min_i x_i} \cdot M \right\rceil & \text{otherwise} \end{cases} \tag{4}$$

The x_{i-1} are processed analogously. The discretization is no problem and has only a minor impact on the accuracy of the HMMM, since M can be chosen adequately at increased compuational costs and storage costs. In this context, the results from psychophysics can also be used for the discretization. As a key concept of the Weber-Fechner law, the "just noticeable differences" as integral part of the human sensory system were identified. This means a user notices a difference in perception ΔP only if the physical stimulus changes S for more than a constant proportion of its actual magnitude, i.e. $\Delta P = k\frac{\Delta S}{S}$. Thus, the hidden states can be defined such that these just-noticeable differences are considered.

With the definition of the hidden states $\mathbf{H} = \{H_{11}, H_{12}, \ldots, H_{MM}\}$, the generation of the transition probablity matrix $\mathbf{P} = \{p_{ij,kl}\}$ is straightforward and can now be estimated using empirical transition probabilities, where $p_{ij,kl}$ is the probability for a state transition from H_{ij} to H_{kl}. In the same way, the emission probability matrix $\mathbf{E} = \{e_{ij,u}\}$ is obtained where $e_{ij,u}$ is the probability that in state H_{ij} the emission, i.e. the user rating, is u. In the case of the ACR scale, it is $u \in \{1, \ldots, 5\}$.

An evaluation of the HMMM for online test #1 and online test #2 is shown in Figure 7 which shows the average user rating for a given sequence of web pages. It can be seen that the HMMM leads to a very good match with a mean absolute relative error $\delta_1 = 4.49\,\%$ and $\delta_2 = 6.24\,\%$, respectively. As a result, the HMMM allows accurate Web QoE modeling which takes into account the memory effect by using two-dimensional hidden system states. Hereby, additional KIFs like the type of user and the type of website can be easily taken into account by integrating them in the emission probability matrix.

VI. Conclusions and Outlook

This paper has introduced the memory effect to the field of Web QoE modeling. Motivated by the fact that a person's current experience of service quality is shaped by past experiences, we investigated the memory effect in the context of three web browsing user studies. The test scenario used considers a user sequentially browsing the pages of an online photo album. Our results show that, although the current QoS level clearly determines resulting end-user quality ratings, there is also a visible influence of the quality levels experienced in previous test conditions. Statistical analysis confirmed the significance of the page load time and the memory effect as key influence factors on browsing QoE. In particular, we found that in addition to the current QoS level the user experienced quality of the last downloaded web page has to be taken into account.

The implications of our results on Web QoE assessment and modeling are twofold: firstly, the design of dedicated Web QoE studies are required to quantify the impact of the memory effect on QoE (cf. III). Secondly, time-dynamics and the internal state of the user (that both manifest in the memory effect) are essential components of the web experience and thus need to be adequately reflected in Web QoE models.

As concerns the latter, we discussed three modeling approaches: Support Vector Machines, Iterative Exponential Regression and Hidden Memory Markov Models. The weighting factors of the SVM indicate that the strength of the memory effect lies in the same order of magnitude as the influence of the PLT. However, the SVM is unable to calculate a MOS score because it treats QoE only as a two-class problem (satisfied vs. dissatisfied). In contrast, the IERMo allows to iteratively compute the MOS for a given sequence of PLTs. The memory effect is taken into account with an exponential decay, as the memory effect diminishes if the user experiences the same QoS level for several web pages in a row. The HMMM also computes a MOS score and takes the memory effects into account via two-dimensional hidden system states and the corresponding emission probabilities. In contrast to the IERMo, the HMMM is able to predict the QoE value of an *individual* user and requires only the state transition probability matrix as well as the emission probability matrix. Thus, the HMMM can be easily used for example in simulations to derive the Web QoE.

We are convinced that this work represents an important first step towards a more accurate assessment and modeling of Web QoE that takes the time-dynamics of human perception into account. Building on these results, we foresee future work to address this challenge along the following two trajectories: firstly, additional web browsing studies need to be conducted to investigate the memory effect in the context of the flow of a true web *surfing* experience, beyond the single photo page. To this end, more sophisticated content such as news and web 2.0 pages as well as longer test conditions that span flows of pages are required in order to expose subjects to more complex but realistic stimuli. Secondly, beyond web browsing, the time-dynamics of QoE perception need to be studied across the whole spectrum of interactive data services, including e-mail, file downloads, as well as progressive downloads that are common for online video services such as YouTube.

References

[1] Nokia, "Quality of experience (QoE) of mobile services: Can it be measured and improved?" White paper, 2004.

[2] A. Bouch, A. Kuchinsky, and N. Bhatti, "Quality is in the eye of the beholder: meeting users' requirements for internet quality of service," in *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2000, pp. 297–304.

[3] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband internet traffic," in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, ser. IMC '09. New York, NY, USA: ACM, 2009, pp. 90–102. [Online]. Available: http://doi.acm.org/10.1145/1644893.1644904

[4] International Telecommunication Union, "Estimating end-to-end performance in ip networks for data applications," *ITU-T Recommendation G.1030*, November 2005.

[5] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, "The Logarithmic Nature of QoE and the Role of the Weber-Fechner Law in QoE Assessment," in *Proceedings of the 2010 IEEE International Conference on Communications*, may. 2010, pp. 1 –5.

[6] J. Shaikh, M. Fiedler, and D. Collange, "Quality of Experience from user and network perspectives," *Annals of Telecommunications*, vol. 65, pp. 47–57, 2010, 10.1007/s12243-009-0142-x. [Online]. Available: http://dx.doi.org/10.1007/s12243-009-0142-x

[7] J. Nielsen, *Usability Engineering*. San Francisco, California: Morgan Kaufmann Publishers, October 1994.

[8] Y. X. Skadberg and J. R. Kimmel, "Visitors' flow experience while browsing a Web site: its measurement, contributing factors and consequences," *Computers in Human Behavior*, vol. 20, pp. 403–422, 2004.

[9] I. C. Jonathan, J. Lazar, K. Bessiere, J. Robinson, and B. Shneiderman, "Determining causes and severity of end-user frustration," *International Journal of Human-Computer Interaction*, vol. 17, pp. 333–356, 2002.

[10] International Telecommunication Union, "Methods for subjective determination of transmission quality," *ITU-T Recommendation P.800*, Aug. 1996.

[11] E. Ibarrola, F. liberal, I. Taboada, and R. Ortega, "Web QoE Evaluation in Multi-agent Networks: Validation of ITU-T G.1030," in *ICAS '09: Proceedings of the 2009 Fifth International Conference on Autonomic and Autonomous Systems*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 289–294.

[12] M. Andrews, J. Cao, and J. McGowan, "Measuring human satisfaction in data networks," in *INFOCOM*. IEEE, 2006.

[13] L. Gros and N. Chateau, "Instantaneous and overall judgements for time-varying speech quality: Assessments and relationships," *Acta Acustica united with Acustica*, vol. 87, pp. 367–377(11), May/June 2001. [Online]. Available: http://www.ingentaconnect.com/content/dav/aaua/2001/00000087/00000003/art00009

[14] B. Weiss, S. Moeller, A. Raake, J. Berger, and R. Ullmann, "Modeling call quality for time-varying transmission characteristics using simulated conversational structures," *Acta Acustica united with Acustica*, vol. 95, pp. 1140–1151(12), 2009.

[15] A. Baddeley, "Working memory: Looking back and looking forward," *NATURE REVIEWS NEUROSCIENCE*, vol. 4, no. 10, pp. 829–839, 2003.

[16] D. C. Rubin, S. Hinton, and A. Wenzel, "The precise time course of retention," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 25, no. 5, pp. 1161 – 1176, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/B6X09-476VXND-4/2/1e79c2a7be40bd2af84ba1fa1c97f90f

[17] A. Luchins, "Primacy-recency in inmpression formation," in *The Order of Presentation in Persuasion*, C. I. Hovland, Ed. New Haven: Yale University Press, 1957.

[18] S. Voran, "A Basic Experiment on Time-Varying Speech Quality," in *Proceedings of the 4th International MESAQIN (Measurement of Speech and Audio Quality in Networks) Conference*. Prague, Czech Republic: MESAQIN, June 2005.

[19] D. S. Hands and S. E. Avons, "Recency and duration neglect in subjective assessment of television picture quality," *Applied Cognitive Psychology*, vol. 15, no. 6, pp. 639–657, 2001. [Online]. Available: http://dx.doi.org/10.1002/acp.731

[20] A. Raake, "Short- and Long-Term Packet Loss Behavior: Towards Speech Quality Prediction for Arbitrary Loss Distributions," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 6, pp. 1957 –1968, nov. 2006.

[21] Z. Chen and H. Nakazato, "Model for time-varying quality of speech services," in *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, vol. 1, nov.-2 dec. 2005, p. 5 pp.

[22] International Telecommunication Union, "BS.1387 : Method for objective measurements of perceived audio quality," *ITU-R Broadcasting Service*, Nov. 2001.

[23] ——, "Continuous evaluation of time varying speech quality," *ITU-T Recommendation P.880*, May 2004.

[24] ——, "Subjective video quality assessment methods for multimedia applications," *ITU-T Recommendation P.910*, April 2008.

[25] ——, "Methodology for the subjective assessment of the quality of television pictures," *ITU-R Recommendation BT.500-11*, 2002.

[26] M. Carson and D. Santay, "NIST Net: a Linux-based network emulation tool," *SIGCOMM Comput. Commun. Rev.*, vol. 33, pp. 111–126, July 2003. [Online]. Available: http://doi.acm.org/10.1145/956993.957007

[27] J. C. Platt, "Using Analytic QP and Sparseness to Speed Training of Support Vector Machines," *Advances in Neural Information Processing Systems*, vol. 11, 1999.

[28] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[29] O. Rose, "A Memory Markov Chain for VBR traffic with strong positive correlations," in *ITC 16*, Edinburgh, GB, 1999, pp. 827–836.