Improving learner quality of experience by content adaptation based on network conditions

Cristina Hava Muntean

School of Informatics, National College of Ireland, Dublin, Ireland

Available online 1 October 2007

Abstract

Apart from user characteristics, properties of the network over which the content is delivered and device on which the content is displayed affect end-user perceived quality. This paper presents a learner quality of experience (QoE) model that apart from the user-related content adaptation, considers delivery performance-based content personalisation in order to improve user experience when interacting with an online learning system.

A comparison-based study on the benefit of using the proposed learner QoE model in adaptive and personalized education was conducted involving the original AHA! and QoEAHA – a version of AHA! enhanced with the learner QoE model. Testing results demonstrate significant benefits in terms of learning achievement, learning performance, learner navigation and user QoE in favour of the learner QoE model-enhanced solution.

© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

The number of people that have access to the internet via different types of wired or wireless networks, the number of information sources that provide different types of content (e.g. Web pages, multimedia, static pictures, etc.) and the number and type of internet-based applications used by users increase exponentially. This growth and diversification led to an increase interest in studying how people perceive their experience
with communication-based systems and various research communities looked at different aspects.

Research in the adaptive hypermedia area for example has demonstrated high benefit in terms of content-related end-user experience when personalized content and navigational support for specific users or user categories is provided. User’s goals, interests, knowledge level, content type preferences, presentation style are some of the human personal characteristics that are currently considered. Lately research has started to take into consideration some performance features such as device capabilities (GUIDE (Smyth & Cotter, 2002), MP³ (Cheverst, Mitchell, & Davies, 2002)), type of the access, communication protocol (SHAAD (Merida, Fabregat, & Matzo, 2002)), state of the network (SHAAD (Merida et al., 2002)), etc. in order to improve further user experience with the personalized service. However, these projects account for only a limited range of factors that may affect the user perception and satisfaction.

Meanwhile, the research in the quality of service (QoS) area seeks to find solutions that bring benefits in terms of QoS during user-system interaction. QoS attributes related to internet-based content delivery services a person perceives include “responsiveness” (i.e. the service availability and timeliness) (Conti, Kumar, Das, & Shirazi, 2002). Service “responsiveness” can be improved by applying:

- **Server-side solutions** – involve admission control mechanisms like requests rejection (request-based (Chen & Mohapatra, 2002) or session-based (Bhoj, Ramanathan, & Singhal, 2000)), content adaptation by degrading the quality of the service provided by all or of some of the clients based on their priority (Abdelzaher & Bhatti, 1999), servers clustering and geographical replication¹ (Cisco Distributed Director), server side caching, etc.
- **Network-based solutions** – include the deployment of network caching appliances (CacheFlow) and “smart” routers and switches (Arrow Point; RouteScience® Adaptive Networking Software) that use on-board intelligence to reduce network latency.
- **Client-side solutions** – include client caching (Kroger, Long, & Mogul, 1997) and pre-fetching (Chen & Zhang, 2003).

A different research direction investigates how users define and perceive QoS and what are user expectations on the QoS delivered by network-based systems. These studies analyse the relationship between objective QoS metrics and the user perception of quality and the impact of certain values of these metrics have on the users. Typically technical measurements based on theoretical mathematical and engineering principles reflect the level of QoS and are removed from measures of end-user perceived quality or Quality of Experience (QoE) (Hestnes, Brooks, Ulseth, Heiestad, & Aaby, 2003). However the latter describes something more than technically oriented QoS and the concept of QoE is attracting more formalised and growing attention especially in relation to Web and multimedia services (Moorsel, 2001; Siller & Woods, 2003).

Although QoE is attracting very much interest, the term has not been generically defined. QoE is considered in Empirix (2001) as a concept comprising all elements of a users perception of the network and performance relative to their expectations. The

¹ http://www.cisilion.com/cisco/cisco-localdirector-.htm.
QoE term is applied to any kind of network-based applications such as Web navigation, multimedia streaming, Voice over IP, etc. According to the application it has different meanings. For example, for a voice over IP application a positive QoE relates to the sound fidelity and ability to smoothly take turns in a conversation. A remote multimedia streaming application has a high QoE if the video image is large and clear when presented to the user. For a Web surfer good QoE means that Web content is retrieved fast enough before getting bored and clicking a link to another Web site. In general QoE is expressed in human terminology rather than technical metrics.

In this context there is a significant need for solutions that would provide users with high QoE in their interaction with various network-based systems. As it is very difficult to provide a generic solution due to the large diversity of existing systems which have various characteristics and requirements, this paper presents results of research that focuses at Web-based systems. A performance-oriented Web user QoE model and learner QoE model – its instantiation in the area of e-learning – are proposed that perform user characteristics-related content adaptation and consider delivery performance-based content personalisation in order to improve user QoE when interacting with Web systems.

Following sections of this paper discuss research attempts to provide certain level of QoE in the WWW area and present means to model user perception of content delivery performance. The proposed learner QoE model is described in details next. A comparison-based study on the benefit of using learner QoE model in adaptive and personalized education involving an original adaptive Web-based system – AHA! and QoEAHA – a learner QoE model enhanced version of AHA! is then presented. Testing results demonstrate significant benefits in terms of learning achievement, learning performance, learner navigation and user QoE in favour of the solution that uses learner QoE model. The paper ends with conclusions and a presentation of future work directions.

2. QoE in World Wide Web

In the area of Web-based adaptive systems that provide personalised information to end-users, research studies tried to relate the objective network service conditions with human perception of Web service quality or user QoE. In the area of World Wide Web, QoE has also been referred as end-to-end QoS or end-user perceived QoS.

Research results (Bhatti, Bouch, & Kuchinsky, 2000; Bouch, Kuchinsky, & Bhatti, 2000; Khirmam & Henriksen, 2002; Krishnamurthy & Wills, 2000) have demonstrated that many QoS parameters such as end-to-end response time (also called page download time or delivery time), network latency, perceived speed of download, successful download completion probability, delivery bandwidth or other defined metrics such as user’s tolerance for delay and frequency of aborted connections factor into user perception of the Web system. Measurement of these parameters may be used to assess the level of user satisfaction. The interpretation of QoS parameters’ values is complex, varying from user to user and also according to the context of the user task.

For example, currently there is a debate on defining what constitutes “acceptable” download time when user experience with a Web site is analysed. It has been acknowledged that long delays cause user frustration leading to performance loss, distraction and difficulty to remember what the user was doing (Shubin & Meehan, 1997). As a consequence, the user may find the Web content less interesting (Ramsay, Barbesi, & Preece, 1998) and of a lower quality (Jacko, Sears, & Borella, 2000). But the question that was
raised is “how long is too long for a user to wait for a Web site response?” Nielsen (2000) suggested a 10 s limit, but recent research (Bouch, Sasse, & DeMeer, 2000; Ceaparu, Lazar, Bessiere, Robinson, & Shneiderman, 2004; Selvidge, 2003) showed that under certain conditions (e.g. often used of the dial-up connection, engaged in completing an important task) users would accept download times significantly longer than 10 s. Willingness to wait more is also moderated by other factors. For instance, novice users and older individuals tend to be willing to wait longer for a computer to react (Schneiderman, 1998; Selvidge, 2003). Users tend to be relatively more patient the first few times they visit a site (Bhatti et al., 2000). Collectively, these findings suggest the users may be more tolerant than the 10-s rule suggested by Nielsen (2000). On average a download time higher then 12 s causes disruption and users start to lose their attention to the task. At the same time it is significant to mention that when the users are aware of their slow connection, they are willing to tolerate an average threshold of 15 s.

Concerning users tolerance for delay in Sevcik (2002) proposed a mapping of the values of the end-to-end response time in human perception space. Three zones of duration that represent how users feel were described: zone of satisfaction, zone of tolerance and zone of frustration. Based on a survey into a number of research studies the authors concluded that users are “satisfied” if the page is loaded in less than 10–12 s. The next zone begins when the page download time exceeds the time limit from the zone of satisfaction. The users start to become aware of the passage of time, slowly building up into annoyance. It is believed that it is a wide band of time between when users are no longer satisfied and when they become frustrated. The zone of frustration starts when the users have reached the point when they are significantly unhappy with the service provided. Values between 30 and 41 s are considered as the critical point (Ramsay, Barbasi, & Preece, 1998; Servidge, 1999; Sevcik, 2002). The conclusion of the research was that 12 s is the upper limit for user satisfaction and 40 s is the limit at which bad performance becomes intolerable.

The effect of two main network QoS parameters on human satisfaction, namely network latency and network delivery speed, was investigated in Khirmam and Henriksen (2002). The authors have concluded that network latency plays a less significant role on the level of user satisfaction when is in the range of 50–500 ms than network bandwidth. However, higher values were not investigated. On the other hand, network bandwidth has a crucial role on the end-user satisfaction when it ranged between 0 and 100 kbps and there is no gain in Web browsing satisfaction for connection speed above 200 kbps.

The frequency of aborted pages, metric defined in Cherkosa, Fu, Tang, and Vahdat (2002), reflects the client satisfaction in relation to the perceived QoS. The main idea behind this metric is that if users get impatient due to a slow perceived access time they will interrupt the transfer of the Web page by clicking either “stop” or “reload” buttons while the page is loading or by clicking a link from the page before the new Web page was loaded. Only aborted pages with a response time higher than a 7 s threshold were considered as reflecting a bad quality of download.

3. Modelling user perception of content delivery performance

Many user-centric personalised content solutions were proposed and developed during the last decade and most of them targeted the educational area. Their aim is to improve user experience by adapting the delivered content to the user interests, goals or skills. However, even if the information is best tailored to the user, the person will not be satisfied
with the on-line information system if the requested content is delivered too slowly or has poor quality. Some important factors that affect user perceived delivery speed and information quality include network-related issues such as connection type, network load, loss and delay, user device-based factors such as screen size, display resolution, processing speed, power levels and most significantly user perception and expectations of user-system interaction process.

Issues related to Web user experience that reflects the overall impression, and satisfaction of the interaction a person has with a Web-based system have been well studied by the HCI community. The Web content delivery-related performance factors and their impact on the end-to-end performance have been analysed by the Web QoS community. Our research builds a bridge between these two research areas by using various results and findings of the two communities. We propose a performance-oriented Web user QoE model that apart from the user-related content adaptation, considers delivery performance-based content personalisation in order to improve user experience when interacting with the system. The proposed model, exemplified and tested in the area of e-learning, is denoted in this paper learner QoE model.

The proposed learner QoE model makes use of a stereotype-based technique in conjunction with probability and distribution theory, in order to construct and infer information about the learner perception of the delivery performance. A number of performance metrics are measured in real-time during the user sessions in order to learn about the current network conditions, and possible changes that may trigger modifications of the learner experience. The learner subjective opinion about the quality of their experience with the system is also used by our model. The goal of the model is to provide suggestions on the properties of the content to be served in order for the learner to have a positive experience with the system that delivers educational content.

4. Learner QoE model

4.1. Stereotype-based structure of the learner QoE model

Stereotypes are widely used in user modelling by systems that provide user personalised information. A stereotype contains assumptions about common characteristics of a group of people and is defined through a list of attributes, attribute values and value estimates.

The proposed learner QoE model consists of a collection of stereotypes $T = (T_1, T_2, \ldots, T_h, \ldots, T_l)$, where a stereotype represents a group of learners with similar perceptions on the performance features. A stereotype $T_h$ has two components: a group of performance features (attributes) $F = (F_1, F_2, \ldots, F_i, \ldots, F_n)$ that characterizes the stereotype and a group of suggestions $S = (S_1, S_2, \ldots, S_j, \ldots, S_m)$ about the properties of the educational content to be served, that would optimise learner perceived performance and ensure a positive e-learner QoE.

Each feature $F_i$ has associated a pre-defined ordered list of linguistic terms (attributes values) $L F_i = (L F_{i1}, L F_{i2}, \ldots, L F_{ik_i}, \ldots, L F_{ip_i})$. Each linguistic term $L F_{ik_i}$ has assigned a numeric value (value estimate) $P F_{ik_i}$, between 0 and 1, representing the probability that the performance feature $F_i$ is equal to the linguistic term $L F_{ik_i}$ for this stereotype $T_h$. (i.e. $P F_{ik_i} = P(F_i = L F_{ik_i} \mid T_h)$). For example, a feature $F_i = \langle \text{learner opinion on the delivery speed} \rangle$ may have associated the following list of pairs LinguisticTerm–Probabilistic-Value: $[\langle \text{bad} \rangle, 0.2], \langle \text{normal} \rangle, 0.5], \langle \text{excellent} \rangle, 0.3]$.
A similar structure is defined for each suggestion $S_j$, which has associated the linguistic terms $LS_j = (LS_{j1}, LS_{j2}, \ldots, LS_{jr_j}, \ldots, LS_{jq_j})$ and probabilistic values $PS_j = (PS_{j1}, PS_{j2}, \ldots, PS_{jr_j}, \ldots, PS_{jq_j})$. For exemplification we consider a suggestion $S_j = \text{"multimedia bitrate"}$ that has associated the following list of pairs LinguisticTerm – ProbabilisticValue: [(1.0,0.3), (1.5,0.6), (2.0,0.1), (2.5,0.0)]. This suggestion indicates that when a multimedia based educational content is delivered to a learner that belongs to the $T_h$ stereotype, the probability that the stream bitrate is 1.00 Mbps or less is 0.3 (30%), the probability that the bitrate is around 1.5 Mbps is 60% and the probability that multimedia can be transmitted at a higher bitrate than 2.0 Mbps is only 10%. Therefore the most suitable bitrate for users in this class should be close to 1.5 Mbps in order to provide a good user QoE in the current network conditions.

A generic representation of a stereotype $T_h$ is presented in Table 1 (group of features) and Table 2 (group of suggestions).

The probabilistic values $PF_{ik_j}$ associated with the group of features are used to determine the degree of match between a learner and a stereotype, whereas $PS_{jr_j}$ values from the group of suggestions indicate the strength of the suggestions. Next, the method used for the computation of the probabilistic values is presented.

### 4.2. Computation of probabilistic values

The Poisson distribution $pois(x,u)$ Eq. (1) for discrete events and its continuous version $pois(x,u)$ Eq. (2) are used for the computation of the probabilistic values associated with the stereotypes’ linguistic terms.

In general, Poisson distribution models a number of events that occur within a given time interval by assigning probabilities. The parameter $u$ controls the shape and is related to the mean and the variance of the distribution in the given time interval. The parameter $x$ identifies a particular event and is represented by an integer: 0, 1, 2, \ldots, $n$

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Group of features for a stereotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature</td>
<td>List (linguistic term–probability)</td>
</tr>
<tr>
<td>F1</td>
<td>(LF11–PF11), (LF12–PF12), \ldots, (LF1r –PF1r), \ldots, (LF1p –PF1p)</td>
</tr>
<tr>
<td>F2</td>
<td>(LF21–PF21), (LF22–PF22), \ldots, (LF2r –PF2r), \ldots, (LF2p –PF2p)</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>Fi</td>
<td>(LFi1–PFi1), (LFi2–PFi2), \ldots, (LFir –PFir), \ldots, (LFip –PFip)</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>Fn</td>
<td>(LFn1–PFN1), (LFn2–PFN2), \ldots, (LFnr –PFNr), \ldots, (LFnp –PFNp)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Group of suggestions for a stereotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suggestion</td>
<td>List (linguistic term–probability)</td>
</tr>
<tr>
<td>S1</td>
<td>(LS11–PS11), (LS12–PS12), \ldots, (LS1r –PS1r), \ldots, (LS1q –PS1q)</td>
</tr>
<tr>
<td>S2</td>
<td>(LS21–PS21), (LS22–PS22), \ldots, (LS2r –PS2r), \ldots, (LS2q –PS2q)</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>Sj</td>
<td>(LSj1–PSj1), (LSj2–PSj2), \ldots, (LSjr –PSjr), \ldots, (LSjq –PSjq)</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>Sm</td>
<td>(LSm1–PSm1), (LSm2–PSm2), \ldots, (LSmr –PSmr), \ldots, (LSmq –PSmq)</td>
</tr>
</tbody>
</table>
If it is defined $\text{gamma}(x+1) = x!$, a continuous version of the Poisson distribution can be written as in Eq. (2), where $x$ is a positive real number

$$\text{poisd}(x, u) = \frac{\exp(x \log(u)) \cdot \exp(-u)}{\text{gamma}(x+1)}.$$  \hspace{1cm} (2)

**Fig. 1** plots $\text{pois}(x, u)$ function for different mean values $u = 1, 3, 5, 7, 9, 11$. Analysing the shape of the Poisson function, one can notice that for $u = 7$ the Poisson function has almost a normal distribution across the $[0, 15]$ interval with a maximum value of 0.15 close to the middle of the interval ($\text{pois}(7, 7) = 0.15$) and the minimum values (close to zero) for $x = 0$ and $x = 15$. In consequence, the interval $[0, 15]$ was considered for the computation of the probabilistic values when using the Poisson function.

As already mentioned the learner QoE model consists of a collection of “l” stereotypes. Probabilistic values for the $T_{l/2}$ stereotype are computed using $\text{pois}(x, 7)$ function that corresponds to a normal distribution. The other stereotypes will have associated probabilistic values computed as in Eqs. (3) and (4):

- for all stereotypes $T_h$, where $1 \leq h < \lfloor l/2 \rfloor$

$$\text{pois}_T(x, u_h) = \frac{\exp(x \log(u_h)) \cdot \exp(-u_h)}{\text{gamma}(x+1)},$$

$$u_h = \frac{15}{2} + (h - 1) \frac{15}{7}; \hspace{1cm} 1 \leq h < \lfloor l/2 \rfloor; \hspace{1cm} (3)$$

- for all stereotypes $T_k$ where $\lceil l/2 \rceil < k \leq l$

$$\text{pois}_T(15 - x, u_k) = \frac{\exp((15-x) \log(u_k)) \cdot \exp(-u_k)}{\text{gamma}(15-x+1)},$$

$$u_k = \frac{15}{2} + (l - k) \frac{15}{7}; \hspace{1cm} \lceil l/2 \rceil < k \leq l. \hspace{1cm} (4)$$

**Fig. 1.** Poisson distributions with different mean values.
For exemplification, it is considered that the model consists of five stereotypes ($l = 5$). The Poisson functions with mean values $u_i$ ($1 \leq i \leq l$) that will be used for each stereotype are illustrated in Fig. 2. The $u_i$ values are obtained by dividing the interval [0, 15] in five equal segments and computing their middle value. This computation exemplifies Eqs. (3) and (4) for $l = 5$. It can be observed that for $u_k$ values closer to zero the peak value of Poisson function increases. This increase is due to the behaviour of the Poisson function.

As the functions to be used for all stereotypes are known, the next step is to compute the probabilistic values that are associated with the linguistic terms using those functions. We consider a list of linguistic terms of length $q$ and the Poisson function $\text{pois}_T(x, u_h)$ associated with the stereotype $T_h$. The probabilistic values ($PF_{ij}$, where $j = 1, \ldots, q$) are computed as follows:

$$PF_{ij} = \text{avg}(\text{pois}_{T_h}(x_j, u_h);$$

$$x_j \in [\text{step} \ast (j - 1), \text{step} \ast j])$$

$$\text{step} = \frac{[15]}{q}$$

Fig. 3 presents an exemplification for a feature $Fi$ from the stereotype $T3$ that has associated $q = 5$ linguistic terms. As the exemplified model consists of five stereotypes, $\text{pois}(x, 7)$ function is used for the computation of the probabilistic values assigned to the linguistic terms. Table 3 presents the computed probabilistic values according to Eq. (5). These values are also plotted in Fig. 3. This computation mechanism is repeated for each feature $Fi$, from the stereotype $T3$ as well as for all other stereotypes.

The same computation mechanism is also applied for the probabilistic values associated with the linguistic terms of a suggestion $Si$, $i = 1, m$ from a stereotype $Th$. Naturally this process will be applied to all stereotypes.

### 4.3. Learner classification into stereotypes

The goal of the classification process is to determine what stereotype classes a learner belongs to and with what probability. Therefore, a degree of match between learner char-

![Fig. 2. Poisson distribution for five stereotypes.](image-url)
characteristics expressed as a list of performance feature – linguistic term pairs Eq. (6) and each stereotype from the learner QoE model is computed as in Eq. (7)

\[
U = ((F_1, LF_{1k_1}), (F_2, LF_{2k_2}), \ldots, (F_n, LF_{nk_n})),
\]

\[
M(Th) = p(Th|F_1 = LF_{1k_1}, \ldots, F_n = LF_{nk_n}) = p(Th|F_1 = LF_{1k_1}) \ast \ldots \ast p(Th|F_n = LF_{nk_n}).
\]  

Eq. (7) is derived from the probability theory, assuming that all stereotypes contain the same set of features, where \(p(Th|F_i = LF_{ik_i})\) represents the a-priori probabilistic value that feature \(F_i\) from the stereotype \(Th\) is equal with the linguistic term \(LF_{ik_i}\). Each element from equation can be computed separately using the Bayes’ rule as in Eq. (8), where \(p(Th)\) represents the a-priori probability distribution of the stereotypes in the model and \(p(F_i = LF_{ik_i})\) represents the a-priori probability distribution of the values \(LF_{ik_i}\)

\[
p(Th|F_i = LF_{ik_i}) = \frac{p(F_i = LF_{ik_i}|Th) \ast p(Th)}{p(F_i = LF_{ik_i})} = \frac{PF_{ik_i} \ast p(Th)}{p(F_i = LF_{ik_i})}.
\] 

Fig. 3. Probabilistic values associated with \(q = 5\) linguistic terms of a feature \(F_i\) from the stereotype T3.

<table>
<thead>
<tr>
<th>(x) Interval</th>
<th>Avg(pois(_T3,(x,7)))</th>
<th>Normalised values (probabilistic values)</th>
<th>Linguistic terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,3]</td>
<td>0.017</td>
<td>0.053</td>
<td>LF(_{i1})</td>
</tr>
<tr>
<td>[3,6]</td>
<td>0.107</td>
<td>0.272</td>
<td>LF(_{i2})</td>
</tr>
<tr>
<td>[6,9]</td>
<td>0.136</td>
<td>0.419</td>
<td>LF(_{i3})</td>
</tr>
<tr>
<td>[9,12]</td>
<td>0.059</td>
<td>0.229</td>
<td>LF(_{i4})</td>
</tr>
<tr>
<td>[12,15]</td>
<td>0.011</td>
<td>0.034</td>
<td>LF(_{i5})</td>
</tr>
</tbody>
</table>

Currently, it is considered that all the stereotypes from the learner QoE model have the same probability distribution. This means that the probability for a learner to belong to any of the stereotype classes is equal with \(1/\text{NoStereotypes}\). However, the model allows
for different probability distributions to be associated to the stereotypes according to the learner population associated with each stereotype class. This provides higher flexibility and more accurate suggestions, but involves also higher complexity.

Since all the stereotypes contain the same set of features and the same set of linguistic terms for each feature, \( p(F_i = \text{LF}_{ik_i}) \) can be regarded as a normalizing factor, which does not influence the belonging of the user to a stereotype class or another.

As an example we consider a learner \( U \) that experienced a Web page download time of 12 s using a 56 kbps connection speed, with the measured round-trip-time (RTT) of 500 ms and with a quality of perceived performance personally expressed by the user as “normal”. These characteristics led to the user classification by the five stereotype-based learner QoE model as follows: \( M(T1) = 0.01, M(T2) = 45.61, M(T3) = 54.23, M(T4) = 0.15, M(T5) = 0. \)

In conclusion, the computed degrees of match indicate the probabilities with which the learner belongs to each of the stereotypes from the model. These values will be used for the determination of the suggestions related to the content properties.

### 4.4. Determination of content properties suggestions

After a learner was classified in one or more stereotypes and the degree of match with each stereotype was computed, the most suitable suggestions related to the properties of the content to be served need to be determined. These content-related suggestions have to ensure a positive learning experience in given network conditions and good user perception of network performance.

Therefore the group of suggestions associated to the stereotypes the learner belongs to are used. Since the stereotypes’ suggestions are similar, but with different strengths, a weighted merge process of these suggestions is performed. First, for each stereotype the strengths of the suggestions have to be computed taking into account the probability with which the learner belongs to this stereotype. This computation is shown in the following equation:

\[
p(S| = \text{LS}_{ik_i}|T_h) = p(S| = \text{LS}_{ik_i}|T_h) * M(T_h)
\]

Next, the additive formula from probabilistic theory presented in Eq. (10) is used to combine the stereotypes suggestions. In Eq. (10) \( P(E_1) \) is the probability that an event \( E_1 \) occurs, \( P(E_2) \) is the probability that an event \( E_2 \) happens and \( P(E_1 \& E_2) \) represents the probability that both events \( E_1 \) and \( E_2 \) occur. Similarly, the formula can be applied for merging the probabilities in order to determine the probability for the events \( E_1, E_2, \dotsc, E_n \) to appear simultaneously

\[
p(E_1 \& E_2) = p(E_{12}) = p(E_2) + [1 - p(E_2)] * p(E_1), \\
p(E_1 \& E_2 \& E_3) = p(E_{3}) + [1 - p(E_3)] * p(E_{12}), \\
p(E_1 \& E_2 \& \dotsc \& E_n) = p(E_n) + [1 - p(E_n)] * p(E_{1\ldots n-1}).
\]

For the weighted combination of the stereotypes suggestions it was assumed that the stereotypes have independent contributions to the final group of suggestions.

Eq. (11) presents an example of how the computation of the strength with which the suggestion \( S_i \) is equal to the linguistic term \( \text{LS}_{ik_i} \) in case that the user belongs to two stereotypes T1 and T2 is performed
\[ p'(S_i = \text{LSikh}_i | T1) = PS^1ikh_i \]
\[ p'(S_i = \text{LSikh}_i | T2) = PS^2ikh_i \]
\[ p(S_i = \text{LSikh}_i | T1 \& S_i = \text{LSikh}_i | T2) = p'(S_i = \text{LSikh}_i | T2) + [1 - p'(S_i = \text{LSikh}_i | T2)] \cdot p'(S_i = \text{LSikh}_i | T1) \]
\[ = PS^2ikh_i + (1 - PS^2ikh_i) \cdot PS^1ikh_i. \] (11)

For exemplification, the same five stereotype-based learner QoE model is considered with the degrees of learner-stereotype match as follows: \( M(T1) = 0.01, M(T2) = 45.61, M(T3) = 54.23, M(T4) = 0.15, M(T5) = 0 \) (see Section 4.3). After merging the suggestions associated with the four stereotypes the learner has a match with, the final group of content-related suggestions is determined. Table 4 presents the results after normalization. This example considers suggestions related to three Web page properties. For example, \( \text{NoObj} \) gives suggestions with different probabilities related to the number of objects to be embedded in the Web page, \( \text{DimObj} \) indicates the total size of the embedded objects and \( \text{DimPage} \) suggests the size of the html file.

Since the main goal is to maximise learner experience from both content information and perceived performance points of view, the learner QoE model tries to offer as much information as possible while maintaining good perceived performance (i.e. content properties under maximum acceptable limits for delivery). Therefore, for a Web page is important to determine the maximum quantity of information (computed as in Eq. (12) using the values from Table 4) that can be delivered to the learner. Any values for the three Web content suggestions that are smaller or equal than the computed ones, will maintain a good learner perceived performance, which is the goal of the model.

\[
\text{NoObj}_{\text{Max}} = 0.15 \times 3 + 0.37 \times 6 + 0.32 \times 13 + 0.14 \times 18 + 0.02 \times 30 = 9.96 \approx 10
\]
\[
\text{DimObj}_{\text{Max}}(\text{KB}) = 0.15 \times 10 + 0.37 \times 20 + 0.32 \times 50 + 0.14 \times 80 + 0.02 \times 200 = 40.02
\]
\[
\text{DimPage}_{\text{Max}}(\text{KB}) = 0.15 \times 5 + 0.37 \times 10 + 0.32 \times 15 + 0.14 \times 20 + 0.02 \times 30 = 12.67
\] (12)

5. Impact of learner perception of delivery performance on the learning process

The learner QoE model aims at enhancing the e-learning process by personalising the educational material according to both learner perception on the provided quality and current network conditions that may have widely varying performance characteristics (e.g. bandwidth, delay, level of congestion, mobility support) and may affect learner QoE. Therefore, a comparison-based empirical study on the applicability of the learner QoE model in the adaptive and personalized educational area was conducted. The experiment

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Final group of suggestions</strong></td>
</tr>
<tr>
<td><strong>Suggestion</strong></td>
</tr>
<tr>
<td>No. Obj</td>
</tr>
<tr>
<td>Dim. Obj. (KB)</td>
</tr>
<tr>
<td>Dim. Page (KB)</td>
</tr>
</tbody>
</table>
involved the integration of the proposed model with a personalized e-learning system called AHA!, creating a QoE-aware AHA system (QoEAHA). AHA!, an open-source project, was developed at the Eindhoven University of Technology and it is currently used as an adaptive hypermedia courseware application that provides personalized educational material based on learner personal characteristic such as goals, interests, knowledge level and learning style. In this context, the new QoEAHA system allows for both learner profile-based and learner QoE-based content adaptations. QoE-based adaptation is applied when the delivery of the personalised document over a given connectivity environment would not provide a satisfactory QoE for the learner. In this case, adjustments of the characteristics of the personalised document are necessary.

The experimental study reported in this paper analyses how the e-learning process is affected by learner perceptions on both content quality and network performance. Learning outcome (learner achievement), learning performance, learners navigational behaviour and their opinion on the QoE provided by the system were analysed and compared when the QoEAHA and AHA! systems were used by learners in a low bitrate home-like operational environment.

5.1. Experimental Setup

The experiments took place in the Performance Engineering Laboratory, Dublin City University, Ireland and involved sixty-two post-graduate students from Faculty of Engineering and Computing, Dublin City University, randomly divided in two groups. The first group of subjects used QoEAHA, whereas the second group used AHA!. The subjects were not aware of the type of the system they were using. For both groups of subjects the same constant network conditions were emulated between their computers and the two systems (Fig. 4). NISTNET, a network emulator that allows for the emulation of various network conditions characterised by bandwidth, delay, and loss was used to create low bit rate home-like internet environments with network connectivity between 28 kbps and 128 kbps. The test results presented in this paper involved a 56 kbps emulated network environment. Tests that emulated network environments with other transfer rates in the 28–128 kbps range provided similar outcomes as the 56 kbps case.

![Fig. 4. Laboratory network configuration for the experimental study.](image-url)
A task-based scenario that involves an interactive study session on chapter one of the AHA! adaptive tutorial was developed and carried out. No time limitation was imposed on the study period.

At the beginning of the study session the subjects were given a short explanation on the usage of both systems and then they proceeded to perform the following tasks:

**Task 1:** complete an on-line pre-test questionnaire (QoEAHA Evaluation Forms, xxx) with six questions that assesses the subject’s prior knowledge on material to be studied.

**Task 2:** log onto the system and proceed to browse and study the material. Back and forward actions through the studied material were permitted.

**Task 3:** complete a post-test questionnaire (QoEAHA Evaluation Forms, xxx) consisting of fifteen questions at the end of the study period. The questionnaire tests recollection of facts, terms and concepts as well as knowledge level after completing the study.

**Task 4:** answer a usability questionnaire (QoEAHA Evaluation Forms, xxx) consisting of ten questions categorised into navigation, accessibility, presentation, performance and subjective feedback.

Pre- and post-test questionnaires were such devised that consisted of a combination of four different types of test-items, commonly used in education: “Yes–No”, “Forced-Choice”, “Multi-Choice” and “Gap-Filling” test items (Muntean & McManis, 2004). Each test-item type has different degree of difficulty and therefore a different weight in the final score is assigned for a correct answer.

5.2. Learner achievement analysis

Learner achievement is defined as the degree of knowledge accumulation by a person after studying a certain material. It continues to be a widely used barometer for determining the utility and value of e-learning technologies. Learner achievement may be analysed in the form of course grades, pre/post-test scores, or standardized test scores. Pre-test/post-test scores technique was used in our experiment.

An analysis of the pre-test scores (see Table 5) shows that both groups of students had the same prior knowledge on the studied material. The average score was 0.35 and 0.30 (out of a maximum of 10) for the students that used AHA! and QoEAHA, respectively. This fact was confirmed by a two-sample T-Test with a 99% confidence level. Therefore the learner achievement can be assessed by processing and comparing directly the post-test scores of the two groups of subjects, presented in Table 6.

<table>
<thead>
<tr>
<th>Table 5: Pre-test scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean score</td>
</tr>
<tr>
<td>AHA!</td>
</tr>
<tr>
<td>QoEAHA</td>
</tr>
</tbody>
</table>
Following the post-test results analysis, the mean score of the subjects that used QoEAHA was 7.05 and the mean grade of those that used AHA! was 6.70. A two-sample $T$-Test analysis on these mean values does not indicate a significant difference in the final marks of the two groups of users ($z = 0.05$, $t = -0.79$, $t$-critical $= 1.68$, $p(t) = 0.21$). The maximum possible score was 30 points, and the subjects’ achieved scores were normalised in the 0–10 range.

Therefore results indicate no significant difference in the learning outcome of the two groups, regardless of the characteristics of the operational environment and despite the suggestions and modifications of the educational content properties performed by the QoEAHA system.

As the studied Web pages consisted of only text and images, degradation of image quality was required in order to meet learner expectation on the delivery performance. In order to assess the effect of image degradation, an analysis of the students’ achievement on three questions (Q9, Q12, and Q15) from the post-test questionnaire was also performed. These particular questions had different weights in the final score and they required the study of the information presented in some images embedded in the Web pages. The mean value of the subjects’ marks for these three questions was 6.3 for the QoEAHA group and 6.4 for AHA! group, after the marks were normalized in the 0–10 range. A two-sample $T$-Test analysis, with equal variance assumed indicates with a 99% confident level that there is no significant difference in the students’ learning achievement ($t = -0.08$, $t$-critical $= 2.71$, $p(t) = 0.93$, confidence level $z = 0.01$). This result is very important as image quality degradations (up to 34% in size for the 56 kbps connection speed) were applied by the QoEAHA.

In summary, these results indicate that a controlled quality degradation performed by the QoEAHA system did not affect the learning outcome, offering the same learning capabilities as the classic AHA! system.

Next analysis assesses the learning performance of the subjects in terms of how quickly the students succeeded to achieve this learning outcome, when the two systems AHA! and QoEAHA were used.

### 5.3. Performance-based analysis of the learning process

Learning performance refers to how fast a study task takes place. A study task may involve a learning process, searching for a piece of information described in the educational material or memorising information displayed on the computer screen. The most common metric used in the educational area for measuring learning performance is Study Session Time (SST) (Ng, Hall, Maier, & Armstrong, 2002; Weber & Brusilovsky, 2001). Apart of this, other metrics such as Information Processing Time (IPT) and Information Access Time (IAT) were measured in our experimental study. These metrics were analysed and compared for both groups of subjects that used AHA! and QoEAHA systems, respectively.
5.3.1. Study session time (SST)

The task-based scenario described in Section 5.1 requires the learners to perform an interactive study. SST is measured from the moment when the subject logs on the system and proceeds to study, until s/he starts answering the post-test questionnaire.

The distribution of the time taken by the students to accumulate the required information using the either AHA! or QoEAHA systems is presented in Figs. 5 and 6, respectively. One can notice that the very large majority of the students that used QoEAHA (71.4%) performed the task in up to 20 min with a large number of students (42.9%) that allocated between 15 and 20 min for the study period. In comparison, when the AHA! system was used, only 42.8% of the students succeeded to finish the learning task in up to 20 min. It is significant to note that over 71% of the AHA! users needed up to 25 min to finish the same task and most of them (28.6%) required between 20 and 25 min.

Summarising the SST results for the two groups in Table 7, it can be seen that on average, the students that used QoEAHA allocated a lower period of time for the study ($\text{SST}_{\text{Avg}} = 17.77$ min) than the ones that used AHA! ($\text{SST}_{\text{Avg}} = 21.23$ min). This led to 16.27% improvement in the learning performance while still achieving the same level of knowledge at the end of the study period as it was confirmed by the learner achievement analysis.

5.3.2. Information Access Time (IAT) and Information Processing Time (IPT)

Time taken to perform a study session (SST) Eq. (13) is a sum of learner perceived access times (download time) of all Web pages – denoted as information access time
IAT, and the time required by a learner to read the displayed information and to accumulate the knowledge – known also as information processing time (IPT). The values of these two components when the AHA! and QoEAHA systems were used by the learners were analysed and their effect on the learning process was studied. The study session involved access to five Web pages generated by the adaptive personalised AHA! tutorial for a person that does not have any knowledge about the concepts to be taught

\[
SST = \sum_{i=1}^{5} \text{InfoAccessTime}_{\text{Page}_i} + \sum_{j=1}^{5} \text{InfoProcessTime}_{\text{Page}_j}.
\]  

Fig. 7 plots IAT per study session and quantity of delivered data when the personalised educational material was retrieved by AHA! or QoEAHA systems, respectively over links with different connectivity in the 28–128 kbps range. One can observe that for low connection speeds QoEAHA improved the learner perceived access time during the learning process by up to 37% with a reduction in total quantity of data sent by up to 42%. This was due to the fact that QoEAHA performs content property adaptations (e.g. image compression) to match the learner QoE model suggestions in order to ensure positive learner perceived performance.

IPT per page was measured from the moment when the Web page is delivered and displayed on the computer screen until the user sends a request for another page. The results indicate that on average a lower time per page (IPT = 4.31 min) was spent by a student to process the information delivered by the QoEAHA system, in comparison to the case

<table>
<thead>
<tr>
<th></th>
<th>Mean SST (min)</th>
<th>Min SST (min)</th>
<th>Max SST (min)</th>
<th>St dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHA!</td>
<td>21.23</td>
<td>12.95</td>
<td>31.84</td>
<td>5.90</td>
</tr>
<tr>
<td>QoEAHA</td>
<td>17.77</td>
<td>9.37</td>
<td>30.38</td>
<td>5.44</td>
</tr>
</tbody>
</table>

Table 7
Study session time allocated by the students when AHA! and QoEAHA were used

![Fig. 7](image_url)

Fig. 7. Analysis on the information access time and the total quantity of transmitted data per study session versus different connection types when the QoEAHA and AHA! systems are used.
when AHA! was used (IPT = 4.95 min). This observation was confirmed by a data analysis procedure. The Null hypothesis ($H_0$) stated that there is no difference between the means values of the two groups and thus the results of the two groups do not differ significantly. Alternate hypothesis ($H_A$) stated that there is a significant difference between the two groups’ means. By performing a two-sample $T$-Test, assuming unequal variances the Null hypothesis was rejected and the alternative one accepted with a 95% confidence level.

5.4. Learner navigational behaviour during learning process

Number of accesses to a page (NoAccesses/page) performed by learners during an online study session may provide indications on the quality of the learning process and on the learners’ attitude towards their QoE. Any re-visit to a page may indicate that the students were not able to recall the information provided in that page due to the fact that either the quality of the content was poor or the students were not fully focused when studying the content. It may also indicate that the learners are not happy with the perceived content delivery performance and they decide to access another page. Later on when they discover that they need to accumulate the information presented in the previous page, they may decide to access that page again.

In our experimental study we have measured how many times the subjects re-visit the Web pages that presented the educational material during the interactive study session. On average, the students that used the QoEAHA system performed a smaller number of re-visits (NoAccesses/page Avg = 1.40) than those that used the AHA! system (NoAccesses/pageAvg = 1.73). An unpaired two-tail $T$-Test with unequal variance assumed, confirmed with 92% confidence that there is a significant difference in the number of visits to a page measured for the two groups of subjects.

How the content delivery performance of the two systems has affected NoAccesses/page was also investigated. One of the delivered Web pages (P3) included a higher number of embedded images and therefore the total size of the Web page was quite high to be delivered over a low bit rate connection. In consequence, the students that used AHA! system have perceived a download time of 24 s with 56 kbps connectivity. QoEAHA system succeeded to decrease the download time to 14 s (considered acceptable for low speed connections) by performing controlled image quality degradation, and thus reducing the quantity of data to be transmitted (Muntean & McManis, 2004). For this particular page we noticed that on average a higher number of accesses (NoAccesses/P3Avg = 1.76) were performed by the subjects from the AHA group than the students that used QoEAHA (NoAccesses/P3Avg = 1.43) (see Table 8).

The analysis of the variability of NoAccesses/P3 for the two groups of students shows that the variance for Group 1 (QoEAHA) ($\sigma^2 = 0.35$) was lower than the variance for the Group 2 (AHA!) ($\sigma^2 = 0.75$). $F$-Test analysis confirms with a 95% confidence level that Group 1 and Group 2 do not have the same variance and the difference between the two groups’ variances is statistically significant. Therefore, it can be concluded that Group 2 has a higher dispersion than Group 1. Looking at Figs. 8 and 9 it can be noted that a larger number of students (an average of 58%) that used the AHA! system performed more than one access to the page during the study session. At the same time, the majority of students that used QoEAHA (62%) performed only one access to P3. These results show that in general the students from the QoEAHA group succeeded to focus better on the studied material because the page was delivered faster and the learning process was not
interrupted by a long download time. In general, long periods of waiting time for getting access to the material annoys the people and disturbs their concentration on the learning task.

5.5. Learners opinion on QoE

The goal of the online usability questionnaire answered by the students at the end of the study was to measure and compare the usability, effectiveness and performance of the two systems: QoEAHA and AHA!. The questionnaire consisted of ten questions related to system usability aspects and performance issues, with answers on the Likert five-point scale: 1 – “poor”, 2 – “fair”, 3 – “average”, 4 – “good”, 5 – “excellent”. The questions were categorised into: navigation, presentation, subjective feedback, accessibility and user per-
ceived performance. The last two question categories seek to assess the learner QoE. Four questions of the survey relate to these two categories and they assess learner opinion in relation to the overall delivery speed of the system (Q6), the download time of the accessed information in the context of Web browsing experience (Q7), the learner satisfaction in relation to the perceived performance (Q9) and whether the slow access to the content has inhibited them or not (Q5). An analysis of the answers for these four questions is presented in this section. Details on the students’ answers on the other aspects were presented in Muntean and McManis (2004).

It can be noticed from the chart presented in Fig. 10 that the QoEAHA system has provided a better QoE for the learners, improving their satisfaction, which was above the “good” level (4 points) for all four questions. The AHA! system scored just above the “average” level (3 points), significantly lower than QoEAHA. This good performance was obtained in spite of the subjects using slow connection (56 kbps) during the study session and not being explicitly informed about this. A two-sample T-Test analysis on the results of these four questions confirmed that learners’ opinion about their QoE is significantly better for QoEAHA than for AHA!, fact stated with a confidence level above 99%, ($p < 0.01$). On overall, the mean value of learner QoE usability assessment, assuming that the questions were of equal importance was 4.22 for QoEAHA and 3.58 for AHA!. This leads to an improvement of 17.8% when using QoEAHA in comparison with AHA!.

Summarising the answers of all questions from the usability questionnaire when all ten questions were considered of equal importance, the results show that the students considered QoEAHA system (mean value = 4.01) significantly more usable then the AHA! one (mean value = 3.73). These results were also confirmed by the unpaired two-tailed T-Test ($t = 2.44$, $p < 0.03$) with a 97% degree of confidence. This increase of 7.5% in the overall QoEAHA system usability was mainly achieved due to the higher scores obtained in the questions related to end-user QoE.

6. Conclusions

The end-user perceived quality of the interaction with an on-line information system is affected not only by user perception and expectations in relation to the interaction with the
system, but also by network-related issues such as connection type and network load and user device factors such as display resolution and processing speed.

This paper presents a learner QoE model which was proposed in educational area that apart from the user-related content adaptation, considers delivery performance-based content personalisation in order to improve user experience when interacting with the learning system.

A comparison-based study on the benefit of using learner QoE model in adaptive and personalized education was conducted involving the original AHA! and QoEAHA – a version of AHA! enhanced with the learner QoE model. Learning achievement, learning performance, learner navigational behaviour and their QoE after the interaction with the systems were analysed when QoEAHA and AHA! respectively were used for study by two different groups of subjects in the same low bitrate operational environment.

As both groups of students received similar marks on a final evaluation test, it can be concluded that after the deployment of the learner QoE model similar learning outcome was obtained. However significant learning performance improvements in terms of study session time (16.27 % decrease), information procession time per page (13% decrease) and number of revisits to a page (decrease) were obtained when using the learner QoE model. It is noteworthy that there were 66% more students which have finished the study in a given period of time using the QoE enhanced system than those using the original version.

A system usability study showed that students thought that the QoEAHA system provided higher end-user QoE than AHA! as QoE-related questions were answered on average above the “good” level by the QoEAHA subjects and only above “fair” level by the AHA! learners. Questions related to the other aspects of the system such as navigation and presentation achieved similar marks for both systems demonstrating that the QoE enhancements did not affect them.

References


