WEB-QOE UNDER REAL-WORLD DISTRACTIONS: TWO TEST CASES

Dennis Guse[1]*, Sebastian Egger*[2]*, Alexander Raake*[3]*, Sebastian Möller*[1]

[1] Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin, Germany
[2] Innovation Systems Department, AIT Austrian Institute of Technology GmbH, Vienna
[3] Assessment of IP-based Applications, Telekom Innovation Laboratories, TU Berlin, Germany
Dennis.Guse@telekom.de, Alexander.Raake@telekom.de,
Sebastian.Egger@ait.ac.at, Sebastian.Moeller@telekom.de

ABSTRACT

Real-world usage of web browsing differs considerably from typically employed laboratory tests on single or multiple page views. Especially when using mobile devices such as smartphones or tablets, several sources of distraction are prevalent while browsing. In the mobile scenario, users are exposed to distracting factors like other people, traffic, announcements etc., whereas at home parallel use of TV or radio services might distract the user. To assess the impact due to context and distractions, this paper presents two studies of web browsing Quality of Experience. The studied factors of possible influence on QoE include one specific browsing task, the environment and a dedicated distraction task. The results show that, in contrast to ratings for multimedia sessions containing audio, video or speech, QoE ratings for web browsing are not affected by the considered contexts or distractions. However, it was found that the primary task for the browsing session had a significant influence on the QoE ratings regardless of the environment or distracting task.

Index Terms— Subjective Quality, User Study, Distraction, Web browsing, Load times

1. INTRODUCTION

In the past, the performance of web browsing and the underlying network was studied in terms of performance indicators related with Quality of Service (QoS) such as throughput, loss rates and loading times. In particular, current approaches to performance monitoring assess the time it takes until a webpage is being fully loaded [1]. This so-called page load time (PLT) has been shown to only partly match the users 'perceived page load time'. Here, the tasks that users undertake are directly related to the perceived loading performance based on the time at which specific task-related elements of the webpage appear on the screen.

Since a number of years, the QoS-related perspective has been complemented by the Quality of Experience (QoE) perspective addressing how users experience multimedia services and technology. Here, the loading of a webpage and the perception of this process by a user has been taken into consideration in QoE research. In large parts of the work to date, the PLT has been used as the parameter to relate with perceived quality (see e.g. [2]). More recently, starting from loading patterns of one page only, the difference between free and task-driven web browsing on QoE has been quantified [3], and first approaches to cover a browsing session that consists of multiple page-loads [4] have been proposed. It could be shown that the time at which relevant elements load plays an important role for web-QoE.

QoE is usually studied under laboratory conditions, that is, providing a controlled environment and repeatable settings, so participants may focus on the presented stimuli or session alone, and therefore give reliable quality judgements. In real-life usage situations however, people typically do not only focus on the stimuli and the quality judgment, but split their mental resources due to environmental factors or parallel tasks, for example triggered by intrinsic or extrinsic factors, or due to specific habits [5]. For example, in mobile situations or in open space offices, people need to react to their environment and distinguish important from unimportant information while working. Another source of distraction might occur due to parallel tasks such as talking on the phone while writing e-mails or surfing the web while watching TV.

Especially, multi-tasking is quite common in entertainment media usage. Brasel and Gips [6] showed in a laboratory study on parallel use of TV and computer that media switching occurs quite frequently. An average of 120 gaze switches between TV and computer occurred with a focus on the computer in 27.5 min. Participants however could not precisely recall their behavior, i.e. they underestimated their media switching frequency.

It has frequently been debated in how far "real" QoE in real-life usage contexts with possibly multiple tasks and environmental distractions may differ from QoE assessed during laboratory tests with single tasks (see e.g. [7, 8]). QoE has been recently distinguished from most quality tests in the lab by referring to these as measuring "Quality (based on experiencing)", in which test participants are focusing on the quality

evaluation task instead of the media usage experience [9]. In this paper, we present our results studying the following research question:

> Does distraction lead to a less critical rating behavior for web browsing quality with different page loading times?

To address this question, we conducted two studies investigating the impact of distraction, i.e. environmental and media multi-tasking, on browsing session quality ratings under delayed loading of the webpages.

The paper is structured as follows: In Sec. 2, we provide an overview of related work on QoE under different tasks and distractions. The setup and detailed results of the two studies are presented in Secs. 3 and 4, respectively. The paper closes with a conclusion.

2. RELATED WORK

The influence of context or concurrent tasks on QoE and human psychophysiological perception in general is not novel. Furnham and Bradley [10] showed that the distraction by background music on primary task performance (here a learning task) is influenced by the personality trait of extroversion. Extrovert persons achieve a higher performance in situations with such distractions than introverts. Strayer and Johnston [11] found that having a telephone conversation while driving is distracting and yields negative driving results. The parallel telephone task leads to a higher reaction time and increases the miss rate of red traffic lights.

In the QoE context, quality frameworks such as [12] have clearly identified and named these influencing factors. In [9] Raake and Egger describe how contextual information and a user's goal do influence the quality formation process. Most of recent QoE research has primarily concentrated on QoE assessment approaches that preclude influence factors for several reasons such as reproducibility and controllability of test conditions etc.

In terms of mobile video quality Jumisko-Pyykkö et al. [8] found that people are less critical under real-life conditions compared to laboratory settings. The studied test environments included public transport as well as sitting in a café. Ketyko et al. [13] showed for mobile video quality that the perceived sound quality depends on the contextual location. In addition, it could be shown that the self-reported focus of a mobile user on the video content is influenced by the number of people around, with a higher number of people leading to lower focus. For speech quality contextual influence factors have also been reported. Gros et al. [7] found that the range of ratings on a quality assessment scale was narrower in a laboratory than an outdoor environment.

Prior work on QoE assessment of web browsing so far has omitted the impact of distraction on quality judgments. Recent work by Strohmeier et al. [3] found a difference in Web-QoE judgments for task-free and task-driven interaction with a webpage. In the latter case, users tend to focus only on the fastest path to solve the task and are therefore more critical against high PLTs that delays their ability to solve the task. In [4], Strohmeier et al. started to investigate QoE over multiple page views in one browsing session, focusing on the impact of varying PLTs over several webpages. An overview on influencing factors on Web-QoE as well as perceptual events have recently been defined in ITU-T Rec. G.1031 [14].

3. EMPIRICAL STUDIES

We conducted two studies using a task-driven Web-QoE assessment methodology as described in [15] and [4] to assess the impact of distraction.

We used the task and system presented by Strohmeier et al. [4] for both studies. The given task is to find the soccer team of the week on a news website with four subordinate webpages starting from the main page. In order to achieve this, a participant has to navigate from the main page to the sports section, then to the soccer subsection and select the article about the team of the week. After clicking on the team of the week article, a dialog box appears and the participant is asked to rate the QoE in terms of loading performance for the web session and to select the correct team of the week. On each of the webpages only one element is active, in the following denoted as the 'relevant element', that is needed for further navigation according to the task, e.g. the sports picture as well as the associated heading 'Sports'. In addition to the QoE rating we also measured the time between the simulated page load starts and the relevant element is selected (click time).

The implementation of the news site is based on the free News Magazine template[1]. The manipulation of the loading process was implemented using in-browser Javascript assigning a load delay to each element individually. The webpage is loaded from a locally running webserver and then the browser is unhiding each element precisely at the pre-defined point in time. The direct instrumentation of load times in the browser made use of a network emulator for manipulation of page loading behavior obsolet. Each of the four pages consists of 11 elements where the load time of each element can be definined for each condition. A screenshot of the main page with an enumeration of the elements is shown in Figure 1.

In Study 1 we investigated passive distraction due to environmental factors focusing on distraction introduced by usage in the public transport system comparable to [8, 16]. In contrast, in Study 2 we introduced active distraction by adding a second task that involves watching a video on TV in parallel to the web-task. The second task also introduces visual distraction as the visual information provided via video must be processed. In both studies we applied a within-person-design

[1] http://www.os-templates.com/free-website-templates/news-magazine

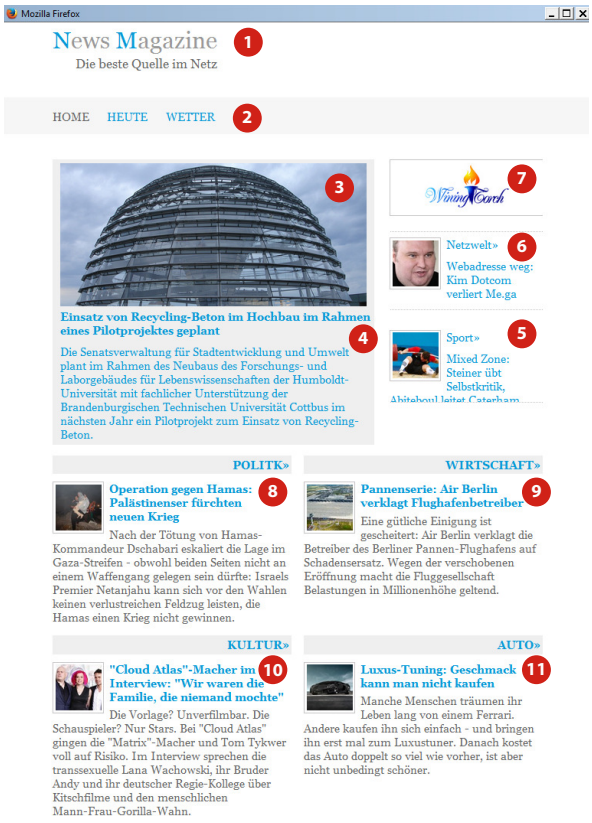

Fig. 1. Screenshot of the main page with enumeration of the 11 elements.

to study the difference between the QoE judgments with and without distraction.

The two studies were conducted using a Microsoft Surface Pro tablet with a 10.6" Full-HD display using touchscreen input.

3.1. Study 1: Web-QoE in public transport

For Study 1 we used two common use cases of web browsing: browsing at home on the sofa, and browsing the web while commuting with public transport. For both use cases the use of a tablet computer is assumed to reflect actual user behavior. The technical solution of a local web server running on the tablet allowed us to use an actual metro line in Berlin for the commuting scenario. In this scenario the test supervisor accompanied the subject and ensured smooth progress of the commuting journey. In contrast, we used our lab facilities for the sofa scenario where the subjects conducted the test alone in the room. Our assumption was that the highly volatile environment in the metro would introduce a certain level of distraction for the subject.

In this study the 11 elements to load were distributed equidistantly over the complete PLT following [4]. In terms of the element loading times we used two different temporal positions of the relevant element. For the 'regular' conditions the relevant element appeared on the 5th position, whereas for the 'delayed' conditions the relevant element appeared on the 9th position (out of 11 elements, cf. Figure 2). Throughout the four consecutive page views of one task the temporal position of the relevant element was kept constant. The PLT used in this study were 2 s, 4 s, 8 s. Three web-task were provided for training and 30 for each condition. In this study a 5-point scale was used for Web-QoE ratings.

In total 25 subjects participated in the test. Their age ranged from 19 to 49 years with mean age = 26 and median age = 29. The subjects were recruited with the prerequisite of having smartphone or tablet experience as we intended to minize learning effects due to touch operation.

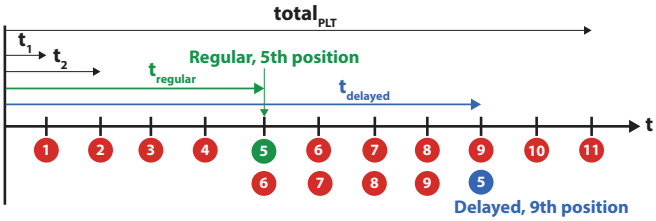

Fig. 2. Position of the relevant element for 'regular' and 'delayed' loading strategies. The numbered circles correspond to the respective elements in Figure 1 and the times t_1,t_2 etc. denote the load times of the respective element.

3.2. Study 2: Web-QoE and television

In Study 2 the distraction due to watching a video parallel to web browsing and the impact on the perception of loading times was studied using the same web-task, system and loading behavior as in Study 1. In addition, a video related task was added that requires giving visual attention to the video: count all visual cuts incl. scene transitions using a handheld counter. Participants were especially instructed to watch the video carefully and not rely on auditory cues for cuts.

The study consisted of two parts, one distraction-free and one with the parallel task, allowing to check for intra-personal differences each consisting of 10 web-tasks. The individual PLT for each web-task was kept constant and was distributed as follows: 2 s, 4 s, 8 s, 12 s, 16 s, 16 s, 12 s, 8 s, 4 s, 2 s. We included the 16 s as we assumed a greater resistance in case of the parallel task. This distribution was selected as it covers a typical situation during usage: a network overload situation occurs and is resolved after some time. In difference to Study 1 the relevant element was shown as either the 8th, 9th or 10th element. In the web-only condition a pause of 10 s was applied between each task whereas for the parallel condition the pause was approximately 2 min.

Study 2 was conducted in a laboratory with a living room ambiente using a Philips Cineos (32", 100Hz 1920x1680,

without Ambilight). As video content we used the movie Océans (France, 2009), which is a documentation about life in the oceans showing underwater as well as terrestrial scenes. The video was cut down to 35 min by removing all arousing or distracting scences, e.g. dying animals, but keeping the original Blu-ray Disc encoding. Overall the study lasted 60 min including introduction, brief training and the two conditions. To avoid order effects of the two conditions the order was randomized.

We used three questionnaires, one to assess the Web-QoE after each task, one after the web-condition and one after the parallel condition. In addition, the NASA Task load Index (TLX) without pairwise comparison [17] was applied after each condition. For the quality ratings a 7-point continuous rating scale [18, p. 19] was used. For this paper the ratings are converted to 5-point scale using [19].

In total 17 subjects participated in this study. Their age ranged from 21 to 39 years with mean and median of age = 26.

4. RESULTS

Within this section we first report on the results for Study 1 and then Study 2 followed by a discussion of the results regarding the research question formulated in Section 1 and future work. Quality ratings are shown as Mean Opinion Score (MOS).

4.1. Study 1

Within this study we varied three independent variables: PLT (ranging from 2 s to 8 s), the position of the relevant element (5th or 9th position) and the distraction level via the context the subject was placed in (sofa and commuting on the metro).

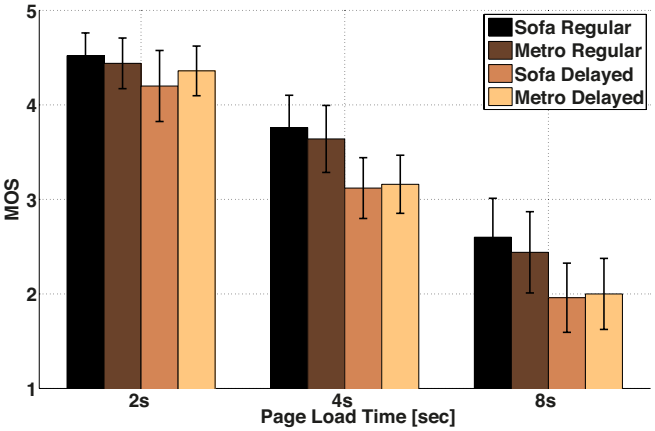

Fig. 3. Study 1: MOS for different PLTs and 'regular' and 'delayed' loading strategies.

The MOS for the independent variables are shown in Figure 3. In terms of different PLTs, significant differences for the QoE ratings can be seen. This is also the case for the different loading strategies, where the 'regular' loading strategy yields significantly better QoE results. However, the hypothesis that subjects are more tolerant towards load times in case they are distracted, which in turn leads to less critical QoE ratings, cannnot be verified within these results. The differences between the sofa conditions and the commuting conditions are not systematically different.

In order to identify if the users were not distracted enough or just had the same perception of load times even with distraction, we evluated the click times. The results are depicted in Figure 4 and show that the subjects did not behave differently in the more distraction context.

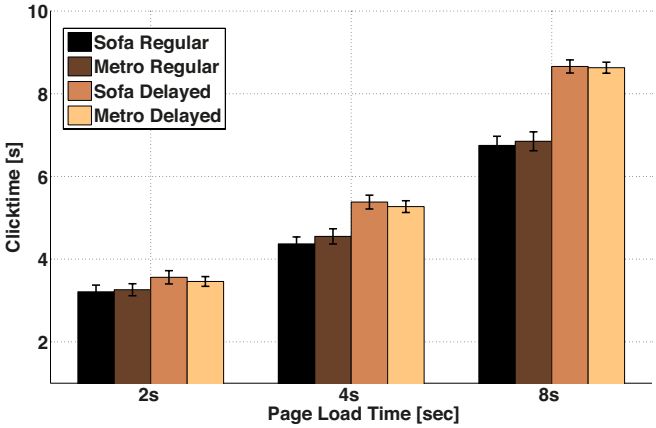

Fig. 4. Study 1: Click times for different PLTs and 'regular' and 'delayed' loading strategies.

Both gathered results failed in identifying differences terms of page load time perception (QoE) and user behavior. Also our qualitative observations of the subject overall behavior while web browsing in the commuting scenario identified no substantial differences between the scenarios. Overall it seems that the commuting scenario did not introduce sufficiently high distraction.

4.2. Study 2

As the results of Study 1 substantiated that the distraction level in the commuting context was not high enough, we opted for a parallel task for this study as described in Section 3.2. In order to investigate the distraction due to the parallel task we applied the TLX. The results in Table 1 show that the mental demand, temporal demand and effort needed for the task completion increase significantly in case the parallel TV task has to be maintained. On the other hand, performance decreases for the condition with the parallel use of the TV, whereas frustration as well as physical demand remain very similar for both conditions.

The QoE rating related results are shown in Figure 5. Besides the 2s PLT condition it can seen that the subjects rated

	Browsing only	Browsing with TV
Mental Demand	15.0 (13.5)	50.9 (19.6)
Physical Demand	8.4 (6.8)	13.8 (13.9)
Temporal Demand	18.4 (20.3)	35.88 (25.8)
Effort	13.8 (12.4)	40.6 (25.4)
Frustration	25.6 (24.1)	25.3 (23.8)
Performance	75.0 (26.6)	62.4 (15.5)

Table 1. NASA TLX results (mean and in brackets standard deviation) for parallel and non-parallel condition.

less critical in case of distraction by the parallel TV task, which supports our hypothesis from Section 2. The contrasting result for the 2s PLT condition is related to the fact that this condition appeared always as first condition in the parallel task conditions. Therefore, the participants were very occupied with the parallel task. Due to this finding we checked for order effects due to the fixed order of the conditions, but except for the 2s PLT none could be found.

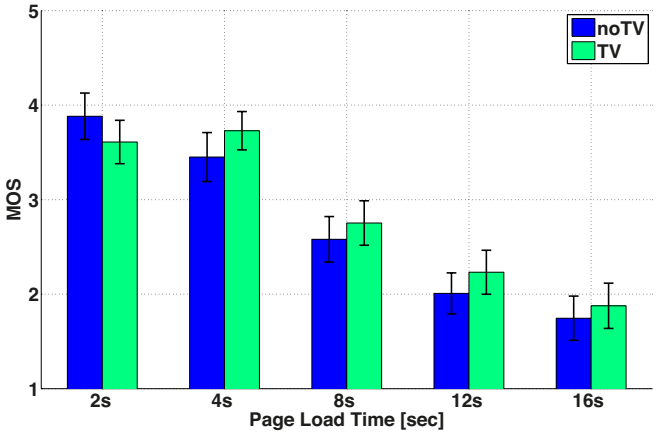

Fig. 5. MOS for different PLTs and with (TV) and without distraction (noTV).

For the click times results where even more pronounced as Figure 6 shows. For all PLTs the click times for the parallel use condition are significantly higher. This proves, additionally to the TLX related results presented above, that the distraction introduced by the parallel task was considerable and influenced the user behavior.

4.3. Discussion

The results presented above show that the distraction levels introduced in the two studies were considerably different. The distraction introduced in the commuting scenario of Study 1 was not severe enough to even influence the subjects clicking behavior. In contrast, the scenario of a parallel task involving a second screen (TV) in Study 2 proved to introduce a considerable distraction. The impact of the introduced distraction

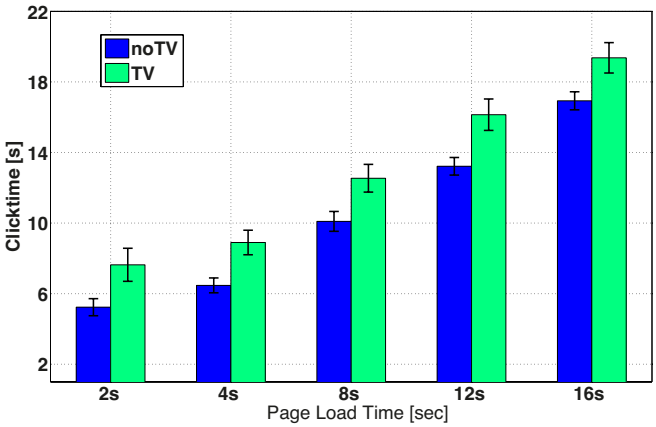

Fig. 6. Click times for different PLTs and with (TV) and without distraction (noTV).

influences the click time behavior and leads to higher TLX scores on mental, temporal and effort. In terms of acquired QoE ratings, Study 1 did not show any significant differences in terms of QoE for the two different conditions, hence not supporting the postulated hypothesis. We believe that the introduced distraction in the commuting scenario is not strong enough to influence the Web-QoE ratings. In contrast, the MOS results of Study 2 identified less critical QoE ratings for the conditions in the parallel use scenario, which supports the postulated hypothesis. Summarising, we acquired results that partially support our hypothesis and found that commuting is not as distracting with regard to Web-QoE as expected. Future work should therefore ensure to quantify the distraction level introduced by the scenarios and gather more QoE data for service usage under distraction.

5. CONCLUSION

Previous work [3] on Web-QoE could show that the task has an impact on the perception of Web-QoE. In this paper we showed with Study 1 that the perception of loading time and browsing behavior is stable against distraction due to environmental factors like crowdedness. By introducing a second simple task, which needs to be solved in parallel, we introduced a more severe distraction source in Study 2. This parallel task influenced the user behavior as expected and also influenced the QoE ratings such that subjects were less critical for these conditions.

In difference to previous results for video streaming, e.g. [8] the environmental factors that caused a better perception of video streaming QoE, are not holding entirely for Web-QoE. In case of video transmission where visual errors can be missed, if visual attention is not given to the playback device, the same does not seem on the same level to hold for the judgment of loading times of a webpage. Further work is necessary to understand the impact of media distraction, as well

as environmental distracting factors on Web-QoE. Especially focusing on the mobile case were the network connectivity is usually more problematic than fixed access connections.

6. REFERENCES

[1] Sebastian Egger, Peter Reichl, T. Hosfeld, and Raimund Schatz, "Time is bandwidth? narrowing the gap between subjective time perception and quality of experience," in *Communications (ICC), 2012 IEEE International Conference on*. pp. 1325–1330, IEEE.

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

[3] Dominik Strohmeier, S. Jumisko-Pyykko, and Alexander Raake, "Toward task-dependent evaluation of web-QoE: free exploration vs.who ate what?," in *Globecom Workshops (GC Wkshps), 2012*. pp. 1309–1313, IEEE.

[4] Dominik Strohmeier, Maija Mikkola, and Alexander Raake, "The importance of task completion times for modeling web-QoE of consecutive web page requests," in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX), 2013*. pp. 38–39, IEEE.

[5] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita, "Habits make smartphone use more pervasive," vol. 16, no. 1, pp. 105–114, 2012.

[6] S. Adam Brasel and James Gips, "Media multitasking behavior: Concurrent television and computer usage," in *Cyberpsychology, Behavior, and Social Networking*, vol. 14, pp. 527–534.

[7] Laetitia Gros, Noël Chateau, and Sylvain Busson, "The impact of real environments on transmitted speech quality judgments," in *Proc. 1st ISCA Tutorial and Research Workshop on Auditory Quality of Systems*, 2003, vol. April 23-25, Akademie Mont-Cenis, D–Herne, pp. 45–50.

[8] Satu Jumisko-Pyykkö and Miska M. Hannuksela, "Does context matter in quality evaluation of mobile television?," in *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. pp. 63–72, ACM.

[9] Alexander Raake and Sebastian Egger, "Quality and Quality of Experience," in *Quality of Experience: Advanced Concepts, Applications and Methods*, Sebastian Möller and Alexander Raake, Eds. Springer, June 2014.

[10] Adrian Furnham and Anna Bradley, "Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts," in *Applied Cognitive Psychology*, vol. 11, pp. 445–455.

[11] David L. Strayer and William A. Johnston, "Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone," in *Psychological science*, vol. 12, pp. 462–466.

[12] Katrien De Moor, Istvan Ketyko, Wout Joseph, Tom Deryckere, Lieven De Marez, Luc Martens, and Gino Verleye, "Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting," *Mobile Networks and Applications*, vol. 15, no. 3, pp. 378–391, 2010.

[13] Istvan Ketyko, Katrien De Moor, Wout Joseph, Luc Martens, and Lieven De Marez, "Performing QoE-measurements in an actual 3G network," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2010 IEEE International Symposium on*. pp. 1–6, IEEE.

[14] International Telecommunication Union, "QoE factors in web browsing," *ITU-T Recommendation G.1031*, December 2013.

[15] International Telecommunication Union, "Subjective Testing Methodology for web browsing," *ITU-T Recommendation P.1501*, 2013.

[16] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti, "Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2005, pp. 919–928, ACM.

[17] Sandra G. Hart and Lowell E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Human mental workload*, vol. 1, no. 3, pp. 139–183, 1988.

[18] International Telecommunication Union, "Subjective quality evaluation of telephone services based on spoken dialogue systems," *ITU-T Recommendation P.851*, November 2013.

[19] Marcel Wältermann, Alexander Raake, and Sebastian Möller, "Comparison between the discrete ACR scale and an extended continuous scale for the quality assessment of transmitted speech," in *Contribution to International Telecommunication Union, ITU-T Study Group 12*, March 2009.