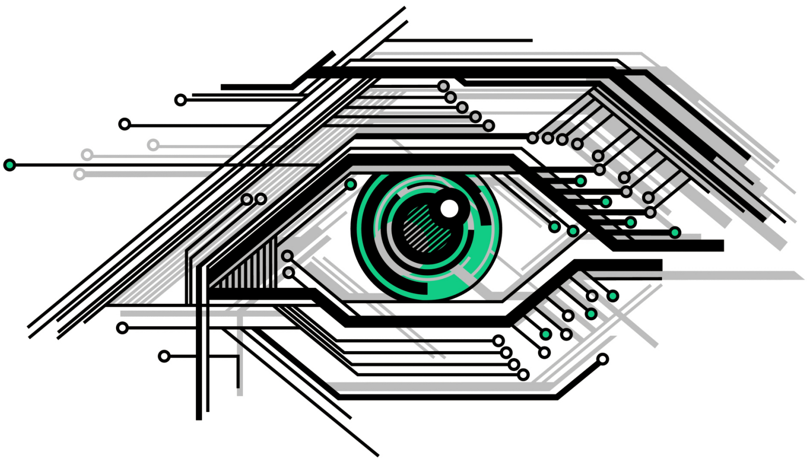
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Concepts for User Experience Research



Note no DART/14/2012

Authors Wolfgang Leister
Ingvar Tjøstheim

Date December 2012

The authors

Dr. rer.nat. Wolfgang Leister, assisting research director at Norsk Regnesentral, received the research doctorate in computer science in 1991 from the Universität Karlsruhe, Germany. His research interests cover smart information systems, multimedia, computer graphics, computer and sensor networks, health care applications, mobile systems, and free software. Recent projects include ADIMUS (Adaptive Internet Multimedia Streaming), DIGEKS (ICT-Based exams), SAMPOS (Strategies for Seamless Deployment of Mobile Patient Monitoring Systems), and CREDO (Modeling and Analysis of Evolutionary Structures for Distributed Services).

Ingvar Tjøstheim is a senior researcher NR. He received the MSc. from the University of Stavanger and Cornell University. He has specialized in user studies, visual technology, market research applications, e-inclusion and accessibility. He has worked closely with the national tourism industry in projects within ICT.

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Concepts for User Experience Research**
Authors	**Wolfgang Leister, Ingvar Tjøstheim**
Quality assurance	Till Halbach Røssvoll
Date	December 2012
Publication number	DART/14/2012

Abstract

We present concepts of user experience (UX) research and show application areas where UX can be applied. We categorise assessment methods, and present estimation models and metrics of UX. Especially the relations between UX and bio-physiological responses, as well as observations are in the foreground.

Keywords	Smart information systems, user experience (UX), Quality of Experience (QoE)
Target group	Researchers
Availability	Open
Project	GB-SMS-2012
Project number	320489
Research field	Smart information systems, User Experience, 3D
Number of pages	37

Contents

1 Introduction

We summarise the work done in the project GB-SMS-2012/13 dealing with *user experience*. The project plan states:

> *Smart information systems shall increase the use, quality and experience of IT applications with respect to costs. In this context, costs can be measured in time, money, resources, work, energy, bandwidth, storage capacity, etc. Improvements in the relation between quality and cost come from research and innovation in the areas technology, system architectures, processes, and value chains. We define metrics and estimation models for objective (measurable) and subjective (personally experienced) quality and experience parameters. We defined the following sub-projects (objectives)*
>
> (*a*) *OCLM (open content lifecycle management) follows up the activity on* collaborative editing*, development of version control for several types of content, relation to licenses for content, and non-destructive editing (Kristoffersen, 2012).*
>
> (*b*) *eHealth deals specifically with communication systems, the Internet of Things (IoT), sensors, and mobile health care applications. We look closer into the concept of the medical digital items (MDI) developed in the SAMPOS project, as well as into semantics in health data, and personal health records (PHR).*
>
> (*c*) *UX (user experience) and QoE (quality of experience) are about subjective and objective assessments and observations, partly relying on physiological data, and applied in systems for access to cultural resources, three-dimensional presentations, eHealth, etc.*
>
> (*d*) *Open data and content are used for IT systems; license compatibility, mobile apps, business models, etc. and is related to the courses INF5780/ITLED4240 at the University of Oslo (Leister and Christophersen, 2012).*

This report addresses Objective (c), i.e., research within UX. The metrics used to measure UX and QoE will be applied to all types of applications with the purpose of adaptivity to offer the best UX in all situations. Application areas span from access to cultural resources, via health care applications (Leister, 2012a) and media presentations to all types of applications where a human-machine interface plays an important part. As a starting point we used the mind map in Figure 1.

Besides this report, the following activities are included in the project:

- Participation in the INREMO network, which is a VERDIKT-funded networking project; this activity includes the preparation of project proposals in the area of UX, together with other INREMO participants.
- Leister (2012b) presented on *Research Challenges for Smart Information System* at the *SMART 2012* conference in Stuttgart.
- Leister and Tjøstheim (2012) presented on *Gamification, User Experience, and Sen-*

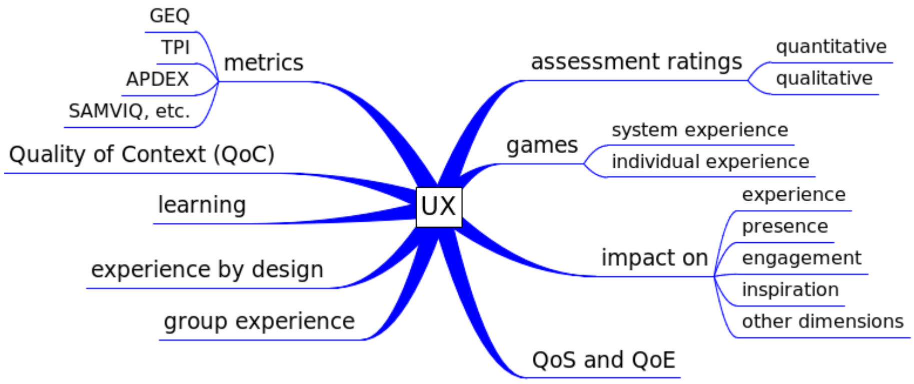


Figure 1. Mind map about UX

sors at the *CRM 2012* conference in Oslo, organised by the DND.

- Participation in an IP proposal to EU's FP7-ICT-2011.8.2: *ICT for access to cultural resources* with the title "Smart technologies, tools, and services for enhanced, engaging and personalised experiences in museums"; this proposal was ranked number four out of 28 IPs.
- Elements of UX research are included in an application to the VERDIKT programme[1], as well as in two applications to the Norwegian Arts Council (Kulturrådet).

2 Smart Information Systems and User Experience

We define *smart information systems* as IT systems that increase quality and experience when using applications and services with respect to costs, as compared to ordinary systems. In this context, costs can be measured in time, money, resources, work, energy, bandwidth, storage capacity, etc. Improvements in the relation between quality and cost come from research and innovation within technology, system architectures, processes and value chains. To evaluate usage, quality, and experience metrics for both, quality and costs need to be defined and applied in assessments including objective and subjective aspects. The rationale to perform such quality assessments is that systems or services that do not show sufficient (subjective) quality or user experience (UX) will be abandoned by the *prosumer*, who is used to choose among many choices.

Smart information systems will have the ability to adapt when measurements suggest that the current UX is decreasing; this is done to keep users satisfied. To be adaptive, smart information systems need to employ sensors that are minimally intrusive; at the same time requirements of privacy, security, etc. need to be fulfilled.

Smart information systems are designed to be used in all application areas, such as eHealth, energy saving and green technologies, and access to cultural resources. Terms

1. This specific VERDIKT call is under the description "smarte IKT-løsninger innenfor helse, omsorg og velferd med brukeren i sentrum"; see the report by Leister (2012a).

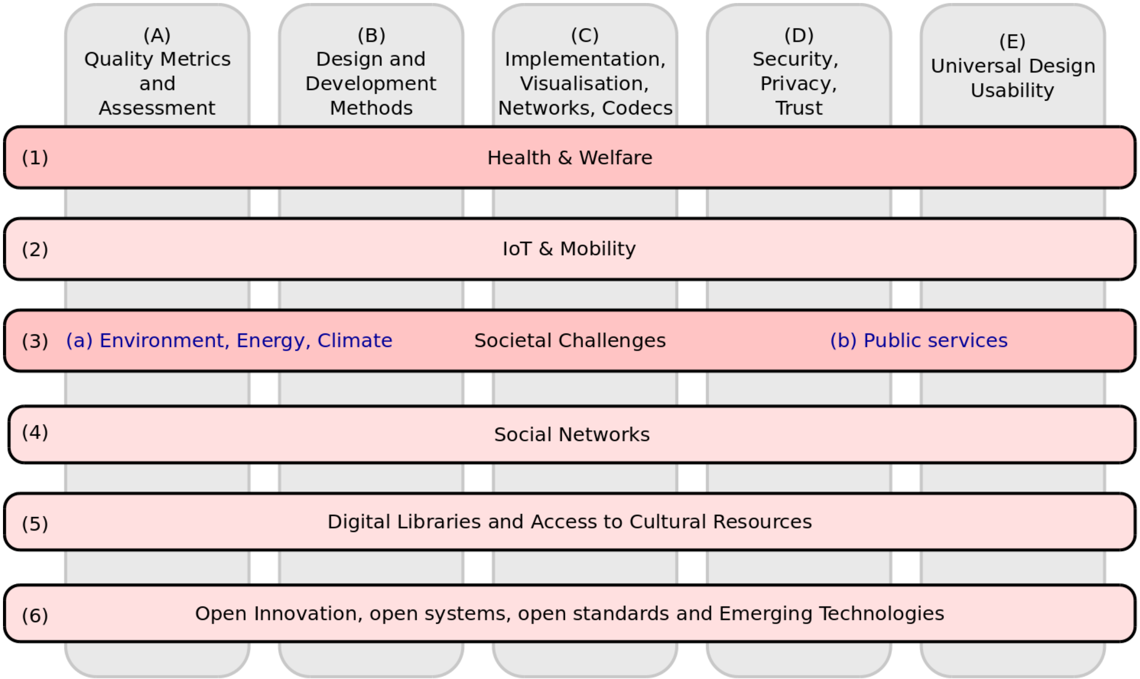


Figure 2. Smart information systems and their context

like smart-house, smart-health, smart-metering, smart-grid, smart-city, and tele-presence are closely related to smart information systems. Figure 2 shows a brief overview of the application areas (horizontal) and the relevant technologies (vertical): *(*A*)* quality metrics and assessment, *(*B*)* design and development methods, *(*C*)* visualisation, networks and codec technologies, *(*D*)* security, privacy, and trust, as well as *(*E*)* usability and universal design.

The term *user experience* (UX) is widely used but understood in many different ways (Roto et al., 2011). Law et al. (2009) give a survey of UX. User experience can be defined as the way a person feels about using a product, system or service.[2] User experience highlights the experiential, affective, meaningful and valuable aspects of human-computer interaction and product ownership. It includes a person's perceptions of the practical aspects such as utility, ease of use and efficiency of the system. User experience is subjective in nature, because it is about an individual's feelings and thoughts about the system. User experience is dynamic, because it changes over time as the circumstances change.

Quality of Experience (QoE) is a subjective measure of a user's experience with a service or system. Without sufficient QoE, a system or service will not be used and eventually be abandoned. Aspects of functionality, usability, universal design, security, etc. are ingredients of such a measure.

Models and metrics for quality exist in other areas: Groven et al. (2012) give an overview of software quality models, and their specialisation to software quality models for Free and Open Source Software (FOSS) (Leister and Christophersen, 2012). Some methods used for assessing the quality of software, such as the OpenBRR (Wassermann et al.,

2. See `http://en.wikipedia.org/wiki/User_experience`.

2005), could be adapted to QoE purposes using different underlying categories and metrics.

Research challenges for QoE include how to apply the subjective and objective QoE/QoS methods to applications; how do these measures have an impact on the social networks; how to collect subjective and objective data simultaneously. We intend to use the most recent eye-tracking glass technology to build a bridge between subjective and objective assessments. Since these eye-tracking glasses also employ a video camera that records the environment, the room and setting of the person, the context can be included in the research design. This also means that data are collected from first-person position, the user views what matters for her. It will give a diversified data collection, allow for method triangulation, and thereby increase robustness of conclusions from the studies.

3 Assessment, Estimation Models, and Metrics

For the assessment of UX and the QoE of an experience several paradigms can be used, depending on the purpose of the assessment. The range of assessments spans from questionnaires to measurement of correlated values and estimation models. We categorise the assessment methods into the following methods:

A) *Subjective assessment:* Subjects are asked to rate an experience. Often, standardised methods are used for certain areas of application. For example, to rate audiovisual systems the MOS rating, SAMVIQ (video) and MUSHRA (audio) are used by individuals. For groups the APDEX or similar indexes can be applied.

B) *Objective assessment of technical parameters:* The assessment is done from technical parameters that are relevant to the human senses. Measures like the peak-signal-noise ratio (PSNR), variances in frame rate, dropouts, etc., but also network parameters such as jitter, delay, etc. can be used. Note, however, that the impact of these measurements to the QoE is not linear.

C) *Estimation models:* The QoE is estimated from objective data that are measured. The estimation models need to be calibrated using subjective assessments before they can be used. Examples of such estimation models include the e-model or methods from our previous work (Leister et al., 2010, 2011, 2007a).

D) *Physiological responses:* The QoE is measured from physiological responses of the subject, such as measuring heart beat, skin conductance (sweat), eye-tracking using specific sensors. Estimation models need to be applied, combining several types of measurements, since there is no linear relationship between the bio-physical measurements and the QoE.

E) *Observation of behaviour and interaction:* From the behaviour of the user, as well as from analysing the interaction of the user with the artifact, conclusions on the subjective experience can be drawn. Such observations can be measurable, such as time used

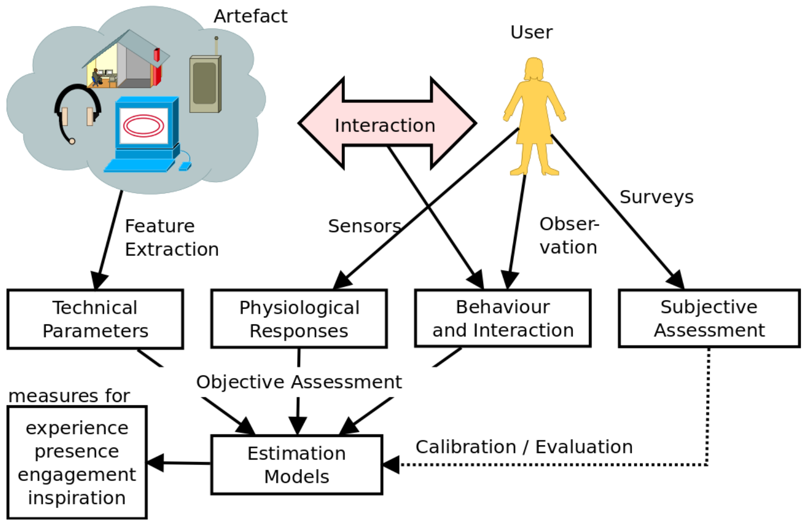


Figure 3. User experience assessment framework

to interact with an artefact, or by applying ethnographic studies, such as behaviour analysis.

F) *Hybrid methods:* The above methods can be combined, e.g., using objective assessment and physiological responses in combined estimation models. Note that subjective assessment is necessary to calibrate estimation models which are part of the hybrid methods.

G) *Groups of individuals:* The above methods A to F can be applied to individuals. For groups of individuals, indices such as variants of the (extended) APDEX can be used, as outlined in Section 3.4. Other metrics consider the interaction between individuals in a group.

Our goal is to develop assessment methods for QoE that are not perceived as being intrusive. Intrusive assessment methods usually reduce the QoE, and, thus, impact the result of the assessment negatively. The assessment should last as long as a subject interacts with an artifact or simply consumes a service, so that the provider of a service around an artifact can use the assessment data to make adjustments in the service level, if necessary. Intrusive assessment methods are usually only applicable in a lab setting. Assessment methods for technical parameters can be classified by whether they are subjective or objective, direct or indirect, in-service or out-of-service, real-time or deferred time, continuous or sampled, intrusive or non-intrusive, and single-ended or double-ended (Tektronix, 2000).

We intend to develop methods where both objective assessment, physiological responses, and estimation models are used in order to derive evidence of how the QoE is perceived for both individuals and groups of subjects in different application areas. To achieve this, we review the assessment methods, correlations between values of assessments, intrusiveness, etc. The goal is to assess experience, but also other metrics such as presence, engagement, inspiration, and, eventually, learning outcome. However, assessing the learning outcome from such assessments is beyond the scope of our work.

3.1 Subjective Assessment

In education science, assessment is the process of measuring knowledge, skills, abilities and attitudes and documenting the same. In general, assessment can be objective (there is a single or multiple specific answers; e.g., multiple choice) or subjective (more than one answer exists; e.g., essay, extended-response).[3] For other types of assessments, which are more relevant in our context, such as the assessment of video quality, audio quality, usability, QoE, etc. a different definition fits better.

For the *subjective assessment* of quality, a pre-established measure or standard is not apparent. Thus, a subjective assessment is based solely on the impression and opinion of the evaluator. The goal of a subjective assessment is being able to tell how the users feel about the object (artifact, software, ...) being tested. This is distinct from how efficiently or effectively individuals (test persons) experience the artifact. The usual method of assessment is to use a standardised opinion questionnaire.[4] Usually, metrics such as the MOS scale or the Likert-scale are used.[5] A Likert scale is a psychometric scale commonly involved in research that employs questionnaires.

For media content, subjective quality assessment involves a test panel with individuals, while the objective measurements are performed on the media content (Pinson and Wolf, 2003a,b). In voice communications and image processing, the mean opinion score (MOS) provides a numerical measure for the QoE using subjective tests (opinionated scores) that are statistically averaged to obtain a quantitative indicator of the system performance. Pre-recorded samples are played back to a mixed group of people under controlled conditions, using the rating: (1) bad; (2) poor; (3) fair; (4) good; (5) excellent. The MOS is the arithmetic mean of all the individual scores.

Since any subjective assessment is based on the opinion of the respective evaluator, answers might be biased due to exterior influences, both physically and psychologically. In order to reduce these exterior influences, the subjective assessment methods define standardised environments. However, this means that the assessment can only be applied in a lab, which implies that the assessment is not performed in the natural environment of the user. In general, lab environments might be intrusive and, thus, have an impact on the assessment result.

3.1.1 Subjective Assessment for Video

Standards for performing user assessment studies to determine QoE for TV and multimedia can be found in DSCQS / BT.500-11 (ITU-R, 2004), SAMVIQ (Subjective Assessment Methodology for VIdeo Quality) / BT.700 (EBU project group B/VIM, 2003; Kozamernik et al., 2004, 2005), and ETR 290, TR 101 290 (Measurement guidelines for DVB systems), and TR 101291 by the EBU. These standards define the methods and conditions concerning how to perform a subjective assessment.

3. See http://wiki.answers.com/Q/What_is_the_difference_between_subjective_assessment_and_objective_assessment.

4. See http://www.usabilitynet.org/tools/subjective.htm.

5. See http://en.wikipedia.org/wiki/Likert_scale.

The subjective assessment methodology SAMVIQ builds on the experiences of the ITU standard BT.500-11 (also known as DSCQS) for video, on experiences from the audio assessment methodology MUSHRA (Multi Stimulus test with Hidden Reference and Anchor), and on the ITU-T recommendation P.911 (Subjective Audiovisual Quality assessment Methods for Multimedia Applications). SAMVIQ has specifically been designed for reliable subjective video quality evaluations, taking into account a range of codec types, image formats, bit rates, temporal resolutions, zooming effects, packet losses, etc. In contrast to TV, multimedia content uses more and different codecs, has variable and non-standardized image formats, temporal frequency, decoder types, display types, and viewing distances. Other differences to the assessment methods include assessment setup, encoding process, viewing panel, training phase, user interface, test organisation, etc.

SAMVIQ considers scenes (visual content), sequences (scene with or without encoding process), algorithms (codec with particular set of encoding parameters), and sessions (part of overall evaluation process). It employs a hidden reference, and uses a subset of all codecs at some rates are used in one session reference.

SAMVIQ is based on random playout of the test files. An individual assessor can start and stop the evaluation process as he wishes, and is allowed to determine his own pace for performing grading, modifying grades, repeating playout when needed, etc. With SAMVIQ, quality evaluation is carried out scene after scene including an explicit reference and a hidden reference. The explicit reference is an uncompressed version of the original sequence and allows the assessor to determine a near-absolute measure of video quality. The hidden reference is technically fully identical to the explicit reference, but is not visible to the subject as such; however the subject should be able to identify it.

The video sequences are shown in multi-stimulus form, so that the assessors can choose the order of sequences to evaluate and correct their votes, as appropriate. As the assessors can directly compare the impaired sequences among themselves and against the reference, they can grade them accordingly.

All the assessors taking part in the evaluation process must be screened in order to establish the consistency of their scores. Inconsistent assessors who produced unstable or even contradictory scores are discarded from the final statistics. In SAMVIQ, all sequences including hidden reference, low anchor and encoded sequences are considered for the screening process.

SAMVIQ is based on the continuous MUSHRA-scale from 0 to 100, with a split of the quality evaluation into five categories. These categories help the subjects to classify their own perception. The subjects shall evaluate their perception of quality according to the following scale: Excellent (80-100), Good (60-79), Fair (40-59), Poor (20-39), and Bad (0-19).

In the course of the MOVIS project, the *Institut für Rundfunktechnik* (IRT) conducted assessments for the WM9 codec using SAMVIQ. For information on the methodology, selection of test sequences, implementation, etc. we refer the reader to the reports by Steinmann et al. (2006); Stoll et al. (2007); and Leister et al. (2007a). This methodology is also summarised in the articles by Leister et al. (2010, 2011).

3.1.2 Subjective Assessment for Audio

For audio, several subjective assessment methods have been developed: PSQM (ITU P.861), PAMS (BT), PEAQ (ITU-R BS.1387), PESQ (Hoene, 2005; ITU Study Group 12, 2001), and the single ended objective measurement algorithm in P.563 (ITU-T, 2005). The triple stimulus with a hidden reference method, as standardised in ITU-R BS.1116, is appropriate for high quality audio. For lower quality levels the MUSHRA is standardised in ITU-R BS.1534.

In PESQ, the quality degradations caused by speech coders and lossy networks are assessed by comparing standard speech samples before and after coding and transmission. The simplifications proposed by Hoene reduce the complexity of the PESQ algorithm to real-time use by using constant length test samples, not time aligning the degraded samples, using loudness levels and not calculating asymmetric distortion using known codecs (ITU Study Group 12, 2001).

The ITU has a newer standard that is able to operate on the received audio only, it is known as a single-ended objective measurement algorithm or P.563 (ITU-T, 2005). The MOS scores of P.563 are more widely spread than those of PESQ, so it is necessary to average the results of multiple tests in order to achieve a stable quality metric. Hence, this approach is not suited for measuring individual calls but can produce reliable results when used over many calls to estimate the service quality.

3.2 Objective Assessment Methods of Technical Parameters/ QoS

Objective assessment methods measure diverse parameters at the technical appliance or application which is part of delivering a service. These parameters are often used to define the Quality of Service (QoS). For audiovisual streaming systems, the user-perceived quality of a service is affected by numerous factors in the end-to-end delivery path, such as encoder, network, decoder, presentation devices, etc.

3.2.1 Assessment of Network Traffic

Existing methods for QoS measurement can be classified into network and application level measurements. Examples of widely used metrics for network-level QoS include connectivity, one- and two-way delay, one-way delay variation (jitter), throughput, and packet loss (Paxson et al., 1998), and active measurement protocols.

Observations of QoS on all levels can be made by non-intrusive measurement (passive observation of QoS at the end system), or intrusive measurement (the controlled injection of content into the stream in order to make deductions of the quality). These measurements are performed according to a QoS metric that quantifies the performance and reliability of the Internet (Paxson et al., 1998; Pezaros, 2005; Smotlacha, 2001).

3.2.2 PSNR

The PSNR (Peak-Signal-to-Noise Ratio) is a double-ended metric, i.e., both source signal and the received signal must be available for calculating the PSNR. The PSNR can be used for different types of media. For video data, the formula $PSNR(\mathrm{dB}) = 10\log_{10}\left(\frac{255^2}{\frac{1}{N}\sum_i (x_i - y_i)^2}\right)$

can be applied (assuming a pixel representation of 8 bit), where the x_i are the samples of the source, and the yi the corresponding samples at the process output. While the PSNR is rather easy to calculate, it addresses only a single aspect, i.e., the noise ratio of the image, while the relevance for a human spectator must be discussed.

Existing work that claims to derive QoE values directly from network QoS frequently applies PSNR. There are very strong arguments against this practice, e.g., Huynh-Thu and Ghanbari (2008); Ni et al. (2009). The picture appraisal rating (PAR) for MPEG-2 (Knee, 2006) assesses the quality of individual pictures instead of videos as well and suffers the same problems as PSNR. A simple approximation in this general family of approaches was proposed by Koumaras et al. (2009). A metric modeling the characteristics of the human visual system (HVS) better is the Structural Similarity, also refered to as $SSIM$ Wang et al. (2004). Compared to $PSNR$, the $SSIM$ is more computationally complex but reflects the properties of the HVS much better with more reliable results.

3.2.3 JND

The just noticable differences (JND) provides an objective metric that tries to emulate the human visual apparatus in a way that considers the change of pictures over time, as do approaches like the Video Quality Metric (VQM) (Wolf and Pinson, 2002). The VQM uses a double-ended approach that can be adapted for continuous in-service quality monitoring when using an additional data channel. VQM is based on the extraction of perception-based features and the computation of particular quality parameters.

3.3 Estimation Models

Estimation models use some type of formula that produces an estimation value from measured values, such as objective measurements. In many cases, subjective assessment is used to calibrate such estimation models. These estimation models can only take a selection of measured values into account. Video quality experts agree that existing quality estimation methods must be used carefully (Mu, 2009).

3.3.1 The R-Value and E-Model

The R-value is a metric for predicting the quality of an audio stream, which includes the effects of delays, jitter, packet losses, etc. in a data network. Different codecs are given a value R. The effects of the data network decrease the original R using impairment factors. See for instance NetPredict Inc. (2004) for an explanation on the effects of codec and network to VoIP. The resulting R can then be computed. Experiments show that the relationship between R and MOS is almost linear.

The E-model of the ITU-T Recommendation G.107 (ITU-T, 2003) uses transmission impairment factors based on a concept given in the description of the so-called OPINE model. The result of a calculation with the E-model is the transmission rating factor R, which is composed of $R = R_0 - I_s - I_d - I_{e-eff} + A$, where all of the factors are composed of several sub-factors. R_0 represents the basic signal-to-noise ratio, including noise sources such as circuit noise and room noise. The factor I_s represents the impairments oc-

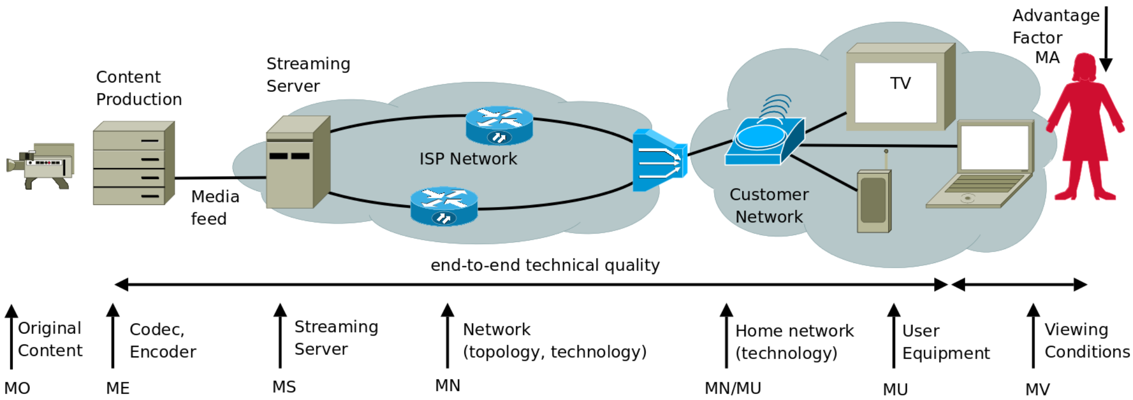


Figure 4. Example delivery chain for streamed content

curring simultaneously with the voice transmission; I_d describes all impairments due to the delay of voice signals, while I_e describes the equipment impairment factor, which is derived from a set of values described in ITU-T Recommendation G.113. The advantage factor A describes psychological aspects under the viewing.

3.3.2 The MOVIS Quality Degradation Model

In the scenario shown in Figure 4, the streamed data are transported from the content provider through various networks to the device where the stream is presented. The MOVIS quality degradation model is outlined by Leister et al. (2010, 2011). Each entity along this delivery chain can potentially mean a decrease in quality as compared to the original quality Q_{O}. We account for this by using the model defined below which estimates the (subjective) QoE for single consumers based on factors reducing the original quality of content.

Inspired by the E-model, we estimate the quality perceived by the end user as a product of the original quality and a number of influencing factors. Each factor is related to a certain entity of the delivery chain in the scenario illustrated in Figure 4 and hereby represents the respective impact on the consumer quality. Thus, the estimated QoE for one consumer is defined as

$$\tilde{Q} = Q_{\mathrm{O}} \cdot \prod_{i \in \{\mathrm{E,S,N,U,V,A}\}} M_i,$$

where Q_{O} is the original quality measure, $M_{\mathrm{A}} \geq 1$, and $0 < M_i \leq 1$ for $i \in \{\mathrm{E,S,N,U,V}\}$. Setting $M_i = 1$ denotes the lack of influence, such as a transparent channel. In the following we describe each single factor:

M_{E}**:** Influence of the encoding on the delivered content. It depends on the codec, codec parameters, and the content (fast vs. slow movements, colour, contrast, etc.).

M_{S}**:** Influence of the streaming server on the delivered content. It depends on the streaming protocol, the implementation of the streaming server and, to a certain extent, on the codec.

$M_{\mathbf{N}}$: Influence of the network on the delivered content. This factor is influenced by technical parameters like delay, jitter, congestion, packet loss, and out of order packet arrival. It also depends on the used codecs and protocols. The parameter consists of three distinct parts: the CP network, the ISP network, and the consumer network.

Possible influences of the hardware at the consumer's home (e.g., routers, WLAN, sharing with other devices) are taken into consideration.

$M_{\mathbf{U}}$: Influence of the consumer's equipment on the delivered content. Hardware type and parameters (e.g., CPU speed, memory size), system and application software, and a system's load parameters have an influence on M_{U}.

$M_{\mathbf{V}}$: Influence of the viewing conditions in the consumer's home, such as distance to the viewing screen.

$M_{\mathbf{A}}$: Advantage factor from the use of the content, modelling cognitive effects like the acceptance of a grainy image sometimes encountered with older films. This increases the value of the content subjectively, even if the technical QoE is worse.

Note that in the general case, some of the factors are not orthogonal, i.e., they depend on the impact factors of previous steps. For instance, a particular networking error can be visible in different manners for different codecs or bandwidth settings.

The factors M_i must be derived from an objective measurement, here called assessment process, and mapped/scaled to the allowable range of values as defined above. Ideally, a model calibration process using regression analysis could be used to derive the scale factors for calculating $\tilde{Q}$. However, this requires a large data set of measurements, so that the dependencies between all input parameters can be derived.

3.4 QoE for Groups of Subjects

The Application Performance Index (APDEX)[6] (Sevcik, 2005) is an open standard for a numerical measure of consumer satisfaction for groups of consumers, for instance for measuring response times. User ratings are categorised into *satisfied*, *tolerating*, and *frustrated*. $\mathrm{APDEX_T}$ is then calculated from the numbers in each category and ranked into five quality classes from *unacceptable* to *excellent*.

Leister et al. (2010, 2011) apply the APDEX based on results from the estimation method outlined. To measure the experience for groups of subjects, media consumers are classified into three quality classes according to their current quality estimate $\tilde{Q}$, given the threshold values T_{S} and T_{U}, like so: The consumer is in the set $M^{(\mathrm{S})}$ for $\tilde{Q} > T_{\mathrm{S}}$, in the set $M^{(\mathrm{T})}$ for $T_{\mathrm{S}} > \tilde{Q} > T_{\mathrm{U}}$, and in the set $M^{(\mathrm{U})}$ else. The threshold T_{S} is suggested to be at 60-80% of the maximum $\tilde{Q}$, and T_{U} at about 40%, depending on the expectations of the consumers. We then apply the formula

$$A_{\mathrm{M}} = \frac{\left|M^{(\mathrm{S})}\right| + \left(\left|M^{(\mathrm{T})}\right| / 2\right)}{\left|M^{(\mathrm{S})}\right| + \left|M^{(\mathrm{T})}\right| + \left|M^{(\mathrm{U})}\right|}$$

6. See apdex.org; accessed August 15, 2011.

and rank A_{M} into U (unacceptable), P (poor), F (fair), G (good) and E (excellent) with the threshold values $0 \leq \{U\} \leq 0.5 < \{P\} \leq 0.7 < \{F\} \leq 0.85 < \{G\} \leq 0.94 < \{E\} \leq 1$. These quality classes are visualised, e.g., using a colour code, for groups of consumers.

There are attempts to extend and generalise the APDEX. Loosley (2010) presents draft versions of the APDEX-G, where several concepts are generalised. APDEX-G defines the *measurement domain* as the set of all possible values to calculate the index from the *Apdex tool*, which calculates and reports APDEX Values. The values from the measurement domain are categorised into *satisfaction levels* (also denoted as *quality levels*, and defined by *performance intervals*. The satisfaction levels of *satisfied*, *tolerating*, and *frustrated* are defined, but it seems that the APDEX-G also can work with a different number of performance intervals. As described above, the APDEX-G is calculated as weighted sum of the members in each satisfaction level, divided by the number of measurements.

3.5 Gamification and Game Experience

Deterding et al. (2011a) define *gamification* as the *use of game design elements in non-game contexts*. To their opinion, gamification refers to *a*) the use (rather than the extension of); *b*) the design (rather than the technology or practices); *c*) elements (rather than full-fledged games); *d*) characteristics for games (rather than play or playfulness); and *e*) in non-game contexts.

The rationale behind gamification is to apply game design thinking, game design techniques, game mechanics and game style to non-game applications to make these more fun and more engaging.[7] Game design techniques and game mechanics are used to solve problems and engage audiences. A key component for gamification to work is engagement. See also the articles by Deterding et al. (2011b,c) for further reading.

Nacke and Lindley (2010) present a literature survey on game experience (GX). They point out that a variety of *models of UX* and *design principles of UX* have emerged with aspects such as emotion, affect, experience, pleasure, hedonic qualities, etc. Since games are artefacts, one needs to distinguish between the *intended* and the *apparent* character of the game experience.

The concept of *playability* is an important aspect of game experience, though inconsistently used in the literature. González Sánchez et al. (2009) propose six *facets of playability*: (*a*) intrinsic, (*b*) mechanical, (*c*) interactive, (*d*) artistic, (*e*) (intra-)personal, and (*f*) inter-personal or social playability. Fabricatore et al. (2002) present a qualitative model of GX, which focuses on utilising the *opinions of players* to develop design guidelines. The literature in game studies uses concepts which are part of a gameplay experience, such as (*a*) flow, (*b*) immersion (*c*) frustration, (*d*) enjoyment, (*e*) presence, (*f*) spatial presence, etc.

According to Nacke and Lindley (2010) there are three methodological categories for experiences that surround digital games: (*1*) the quality of the product (game system expe-

7. See `http://gamification.org/`.

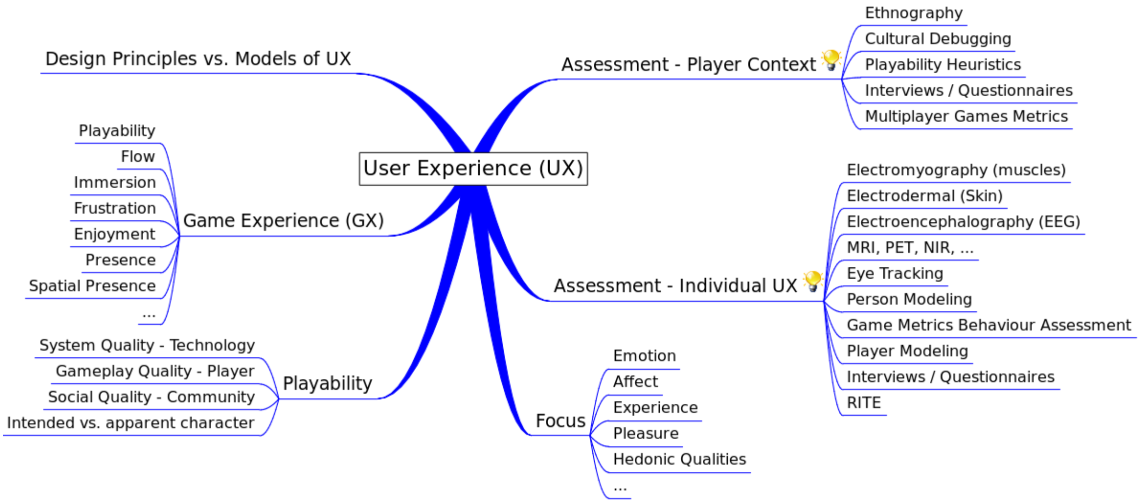


Figure 5. Game Experience map

rience), (*2*) the quality of the human-product interaction (individual player experience), and (*3*) the quality of the human-product interaction in a social, temporal, spatial or other context. The UX evaluation of gameplay can be categorised in three GX frames: (*1*) game experience methods on the functional level of assessing the functional capacity of the game system, game engine, or level data to balance the game design; (*2*) player experience methods to evaluate the emotional or cognitive impact on the player; and (*3*) context experience methods to study the interaction in the light of social impact and player behaviour.

Nacke and Lindley point out that capturing individual emotions using new sensor technology and behavioural tracking can be used to model and assess player cognition and emotion during gameplay. Examples of such technologies include (*a*) psycho-physiological player testing, including electromyography (EMG), electrodermal activity (EDA), electroencephalography (EEG), and magnetic resonance imaging (MRI); (*b*) eye tracking; (*c*) persona modelling; (*d*) game metrics behaviour assessment; (*e*) player modelling; (*f*) qualitative interviews and questionnaires; and (*g*) rapid iterative testing and evaluation (RITE) testing.

For the subjective assessment of UX in games, the *Game Experience Questionnaire* (GEQ) (IJsselsteijn et al., 2007) has been developed. The GEQ contains modules for assessing experiences during and after game play, as well as social presence. The questionnaire captures matters like positive affect, competence, immersion, flow, and challenge. However, some items are difficult to fill in by participants when they only have a short time available to play the game. Nacke et al. (2009) use the GEQ in connection with eye-tracking.

Lombard and Ditton (1997) identified six different conceptualizations of presence: (*1*) presence as social richness, (*2*) realism, (*3*) transportation, (*4*) immersion, (*5*) social actor within the medium, and (*6*) the medium as social actor. They summarise presence as the *perceptual illusion of non-mediation*, i.e., that a person fails to perceive or acknowledge the existence of a medium in his or her communication environment and responds as if the medium would not be there. The Temple Presence Inventory (TPI) (Lombard et al., 2009) defines a measure of presence based on the six conceptualisations above by defining a comprehensive set of questionnaire items. The TPI is used in a variety of contexts (Chatting et al., 2006; Horvath and Lombard, 2010; Tsui et al., 2012). Fukuda (2011) uses the TPI to assess issues of universal design for older adults.

An overview of more assessment methods can be found at `http://www.allaboutux.org/all-methods`.

4 UX Assessment Using Physiological Responses

The subjective assessment models are often considered intrusive, while assessment and estimation models using technical paramenters only indirectly assess the UX by using measurements from technical devices that define the QoE in the production chain of content. In the current section, we review how psycho-physiological responses can be assessed by using sensors, cameras or other devices measuring change of emotion, and how these responses can be used to evaluate UX.

4.1 The Link between Psycho-Physiological Reactions and Emotion

In medicine and in some of the subfields of psychology, biomedical sensor data is used for a range of purposes. Biomedical sensor readings and their relation to emotions have been a research topic during the last two decades (Cacioppo et al., 1993; Cacioppo and Gardner, 1999; Cutmore and James, 2007; Gross and Levenson, 1995; Hatfield et al., 1993; Larsen et al., 2008; Rapson et al., 1993). These studies suggest that emotional reactions can be captured using biomedical sensors. Although there is a correlation between sensor data and emotion, there are some drawbacks related to the fact that the findings are based on laboratory studies. In a laboratory setting, the instruments can be fine-tuned and the researchers have a high degree of control. In a real-world setting, instruments and sensor technology need to be robust and non-obtrusive while being able to provide the intended measurements, only requiring minimal technical support.

The correlation between UX and emotions is closely related to research considering emotion when subjects look at advertisements, and how to capture emotion for market analysis purposes. The findings by Gross and Levenson (1995) in their study about emotion induced by clips from movies document that amusement is an emotion that is distinct, and that can be measured with a self-reported survey instrument. A large number of studies have indicated that affective responses to advertisements have direct effects on the consumers' attitudes toward the ad, and at least indirect effects on the consumers' attitudes toward the brand (Aaker et al., 1986; Batra and Ray, 1986; Brown et al., 1998; Brown and Stayman, 1992; Cohen et al., 2008; Edell and Burke, 1987; Holbrook and Batra, 1987).

According to Wang and Minor (2008) we distinguish fifteen primary categories of human emotions. In marketing, consumer behaviour, and advertisement, satisfaction is fre-

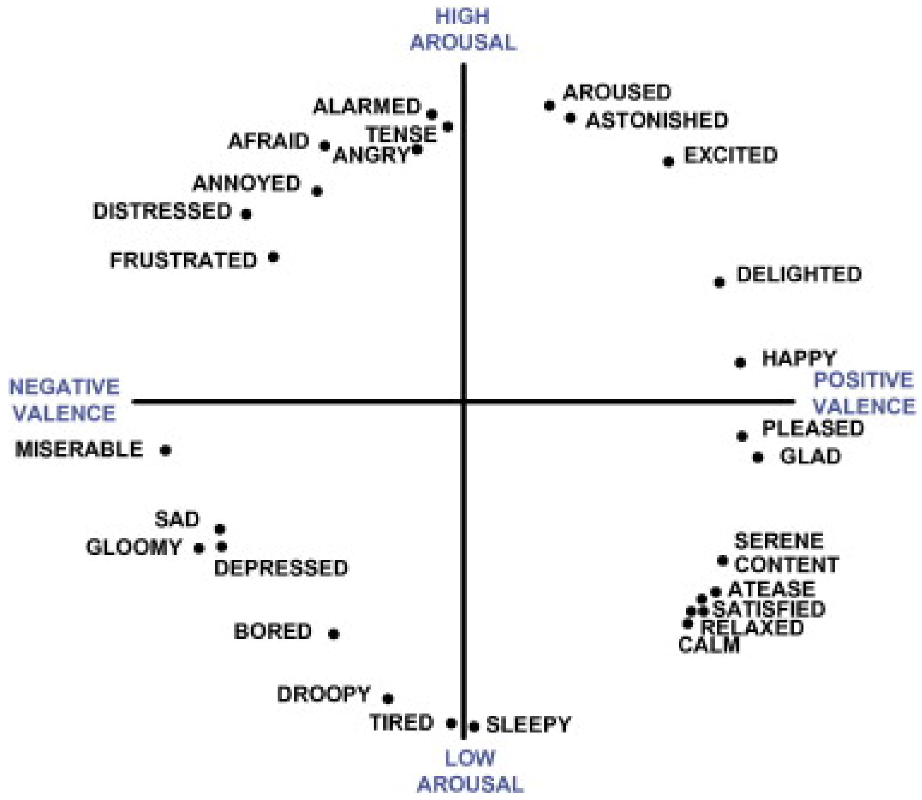
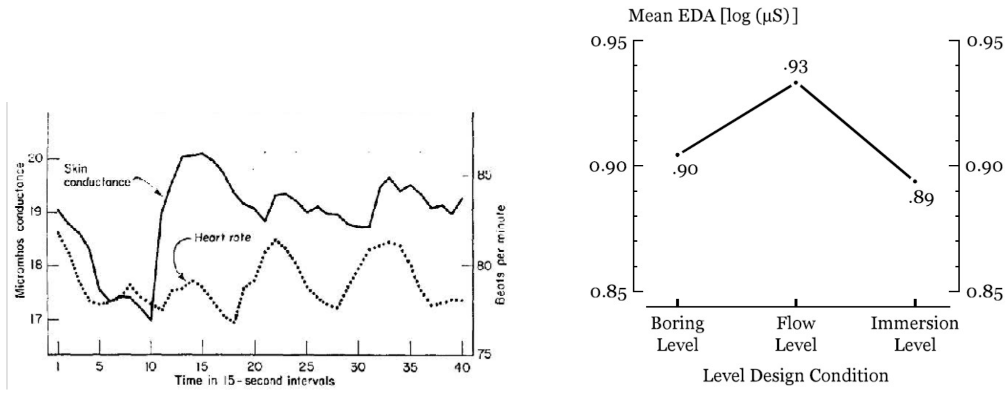


Figure 6. Russell's model of affect with some emotions mapped onto it (Ferreira, 2008; Ibáñez, 2011; Wilhelm and Grossmann, 2010)

quently used. As companies strive for having as many satisfied customers as possible, when a consumer rates a product or service the question is typically "how satisfied are you with ___ (*name of product or brand*)". *Satisfaction* is one of the aforementioned fifteen categories. Two other categories are also particularly relevant for market research: *sensory pleasure* and *excitement*.

Russell (1980) associates two dimensions to emotions – *valence* and *arousal*, as shown in Figure 6. This model indicates that some types of emotion can be inferred from biomedical data. Valence is regarded as an individual dimension. Arousal, however, has been strongly correlated with physiological measures such as the *galvanic skin response* (GSR) (Boucsein, 1992), which measures the electrical conductivity on the skin. Lang (1995) reports a linear correlation between electrical conductivity and arousal, which might enable technologies for a fast physiological indication of arousal.

Skin-conductance, measured in Siemens or mho (Micromho), can be divided in two distinct groups of components: the *tonic* and the *phasic* components. According to Sanches (2008), the tonic component changes slowly over time and is related with the overall arousal level and natural day-to-night variations. The phasic component rides on top of the tonic and causes peaks on the signal when a person is startled, feeling pain, sustaining breathing, etc. These peaks, denoted as *skin conductance responses* (SCR), have been used widely as a measure of instantaneous arousal (Sanches, 2008, p.38).



(a) Skin conductance and heart rate, from Sanches (2008)

(b) Skin conductance (EDA) and flow experience, from Nacke and Lindley (2010)

Figure 7. Skin conductance and emotions

To measure heart activity, the electrocardiogram (ECG) is used. The heartbeat rate has been shown to correlate significantly with skin conductance (Malmstrom et al., 1965). As an example, Figure 7a shows an example where both heart rate and skin conductance were measured. See the thesis by Sanches (2008) for further details. Physiological reactions to emotions have an influence on the ECG, electro-myogram (EMG), galvanic skin resistance (GSR), skin temperature, pulse, etc. from the autonomous nervous system via peripheral signals, whereas signals from the central nervous systems can be measured by, for instance, electro-encephalograms (EEG). It is pointed out that it is possible to distinguish fairly reliably between several discrete emotional states. For a more extensive review of psycho-physiological monitoring, types of bio-signals, and emotion research, see the article by Wilhelm and Grossmann (2010).

4.2 Use of bio-sensor in the studies of video-games

Psycho-physiological measurements can be used to quantify a video-game player's reactions and emotions (Choi et al., 2009). The first person shooter (FPS) game *Half Life* was used in a study with 36 participants by Grimshaw et al. (2008), where they collected both psycho-physiological data, such as GSR data, and questionnaire data. In this study, Grimshaw et al. conclude that the psycho-physiological data did not support the subjective results. They discuss a number of reasons why this was the case, and the need for methodological improvements in order to be able to measure with better precision what they intended to measure. The team of researchers also did a study on the flow experience. The measurement of arousal based on electrodermal activity showed in this case there was a significant difference; the flow level was physically more arousing compared to the two other levels or states (Nacke and Lindley, 2010).

4.3 Context matters

In the review article on affect in consumer behaviour, Cohen et al. (2008) emphasise that information about context is necessary; they state that "the impact of affect in judge-

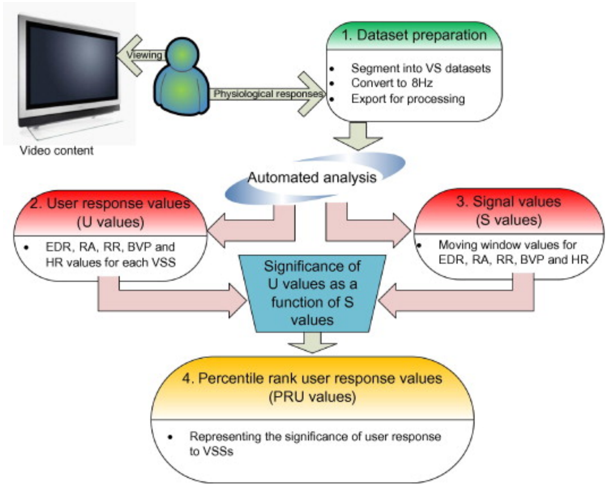


Figure 8. Method used by Money and Agius (2009) to use physiological responses for quality assessment

ment and decision-making cannot be well understood without placing it in a more multifaceted context." (p. 79). A lab study gives the researcher high level of control, but when the goal is to understand human reactions and behaviour in context, it is in the normal (every-day) situation. Schachter and Singer (1962) proposed a two-factor theory for evaluating emotion, suggesting that one needs to know two things to interpret an emotion: (*1*) the physiological arousal from the body, and (*2*) cognition about the context.

There is theoretical support, as well as support from medical and psychological studies for interpreting bio-data as emotional reactions to stimuli. However, outside a laboratory context this interpretation is not easy to extract. Therefore, interpreting bio-data outside a laboratory context might meet obstacles. In a medical context, body-sensors are used by patients based on instructions by medical doctors. Patients or persons in a training context might be willing to use equipment and sensors that are somewhat obtrusive.

Individuals in their home-environment, e.g., watching TV or using a media device might have a lower willingness of wearing or using devices that are perceived as obtrusive. Devices that can be attached to the arm are considered less obtrusive than devices that have to be attached to the hand or to fingers. Therefore, when we discuss wearable devices with biosensors, we need to consider obtrusiveness as a factor.

4.4 Combining Biosensor-Data with In-Situ Surveys

When an individual answers a survey or takes part in studies in which the purpose is to elicit opinions and preferences, we see that context and situation matter. A system that is set up to monitor and intercept the respondents in vivo has several advantages to the traditional survey approach.

Money and Agius (2009) demonstrate a system for video content summarisation using electro-dermal response (EDR), respiration amplitude (RA), respiration rate (RR), blood volume pulse (BVP), and heart rate (HR) as the physiological responses to video content, as illustrated in Figure 8. They only utilise peak (significant) values. In their study, video-

Table 4. The identification of normal/abnormal conditions

Conditions	Body Temperature		Skin Conductance		Heart rate	
	Level(8)	Values	Level(14)	Values	Level(16)	Values
Abnormal	1-4,6-8	otherwise	1-9, 11-14	otherwise	1-11,13-16	otherwise
Normal	5	29-31	10	1700 ~ 1800	12	80~90

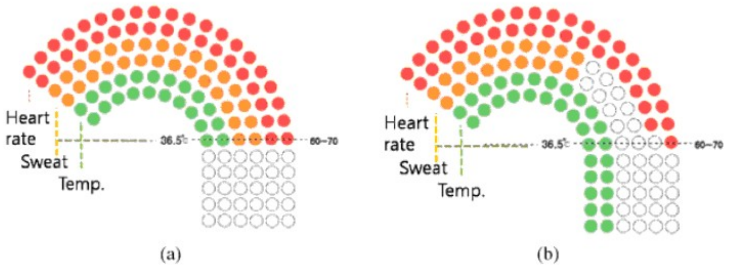


Fig. 14. Physiological information visualization with Rainbow Display (a) normal condition (b) abnormal condition

Figure 9. Visualisation of an emotional response/reaction

content in the categories action/sci-fi and horror/thriller has peaks that can be translated to a (stronger) emotional reaction. A visualisation of normal and abnormal physiological conditions is shown in Figure 9.

4.5 Devices with Biosensor(s)

In many scientific studies, advanced tools have been developed and used in testing hypotheses, such as psycho-physiological reactions in the human body. In many cases, these tests are performed in a controlled environment such as a lab (Cutmore and James, 2007), where lab equipment can be attached to the participants. Devices that are obtrusive and, thus, only can be used in a lab, are not further discussed since our focus is on performing measurements in a normal (in-situ) situation.

4.5.1 The Empath wristwatch

The company Exmovere has developed a wristwatch that currently is called *Empath Watch* (or BT2). It is described by Andrews et al. (2008), and in the document "The new Biotechnological Frontier: The Empath Watch"[8]. Vittorias et al. (2008) show in an example how the Empath Watch can be used by a child with autistic disorders. Here, the authors present an adaptive system that measures the emotional state, such as changes in stress level, hyper-/hypo-tension, during the child's educational procedure. However, Ferreira (2008) has reported after tests that the wristwatch did not work properly.

4.5.2 Devices developed or used by researchers at MIT

At MIT, members of the research group for "affective computing" use a set of devices in their experiments and studies. Examples include a HandWave Bluetooth skin conductance sensor, a galvactivator glove, a GSR mouse, and a wearable EDA sensor. In a recent

8. See http://www.exmovere.com/pdf/Exmovere_Wearable_Sensor_research.pdf.

Figure 10. Picture of the first and second (current) model of the wrist watch

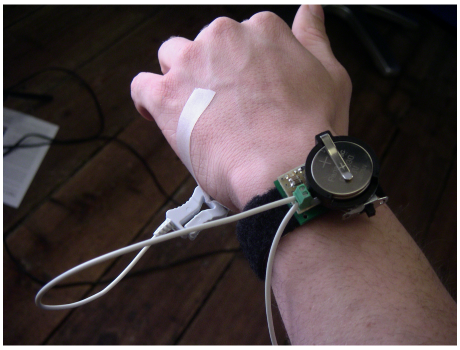


Figure 11. Device developed at MIT (Reynolds, 2005)

paper, Poh et al. (2010) describe the wearable EDA as follows: *" [. . .] the complete wearable EDA sensor is compact (70 mm × 70 mm × 20 mm), lightweight (40.3 g), and the components used can be purchased off the shelf for approximately $150. In contrast, a commercial system such as the Flexcomp Infiniti (Thought Technologies, Ltd.) measures 130 mm× 95 mm × 37 mm, weighs 200 g, and costs $6000 for the data acquisition unit, and an additional $275 for an EDA sensor."*

Picard (2010), professor in the media group at MIT, writes in a review about the mobile electrodermal activity (EDA) and motion sensors used in their research: *"[the sensor] is now being made by Affectiva. Other companies such as BodyMedia, FitSense, and Polar have developed commercially successful heart rate and activity monitoring products for health and fitness. While none of these measures emotional state directly, these devices do capture physiological changes that co-occur with emotional states, providing objective information related to both arousal and valence, which can be interpreted accurately provided that other influencing variables are held constant (e.g., did the temperature, humidity, electrode pressure stay the same and the EDA go up?) While you cannot perfectly control these variables when participants leave the lab, few controls are perfect in the lab either, and significantly more naturalistic data can be obtained leaving the lab. New technologies allow measurement out in the world, comfortably enough that participants forget they're wearing them: These open up whole new areas of inquiry in emotion research."*

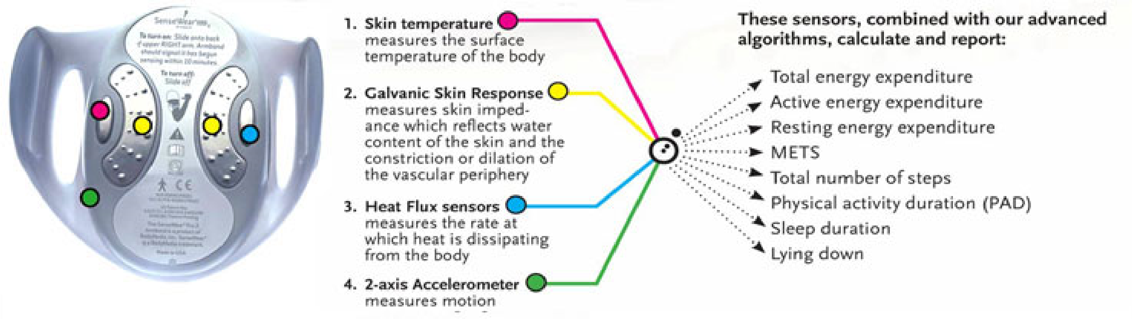


Figure 12. The Sensewear Armband

4.5.3 Sensewear

The sensors in the Sensewear Armband are used to monitor heart beat, heat flux, near body ambient temperature, galvanic skin response, skin temperature, and movements (Liden et al., 2010). It has been used in a number of studies in the telemedicine domain and by consumers who, for instance, intend to monitor weight-loss. According to Tamás et al. (2007), the variation of GSR at the upper arm measure by Sensewear is a good reflection of the evaporative heat loss through the body surface. A quick literature survey shows that the Sensewear Armband has been used in many domains and with good results. It is considered an unobtrusive device, which is also an argument for further using it in new areas.

4.6 Eye tracking and Eye Glasses with Cameras

Recently, the eye-tracking specialists Tobii[9] and SMI[10] launched new wearable products for mobile eye-tracking (Bulling and Gellersen, 2010) in assessment processes. With these types of glasses, a video from a first-person position can be recorded. More important is what a person is looking at, the *gaze points* (in eye tracking terminology).

Eye tracking has in some cases a clear advantage compared to surveys that rely on what a person remembers/recalls. With mobile glasses, eye-tracking of a person on the move is possible. From a methods point of view, valuable video data can be collected together with and as a supplement to the subjective assessments. With this kind of data collection tool, it is possible to link subjective and objective data. Eye-trackers, both stationary and mobile systems, are shown in Figure 13.

Eye glasses that support augmented reality are emerging; these eye glasses, as shown in Figure 14, have a projection system that allows overlay of the normal sight, and additional content projected into the field of sight. While these systems earlier were available as head-mounted displays (HMD), the new prototypes of such glasses are more lightweight and thus less intrusive. When combining these glasses with built-in cameras, an information system can observe the environment of users and what they are potentially observing. It is also possible to implement cameras that observe the eyes of the user, thus making eye-tracking combined with augmented reality features feasible. Therefore,

9. See http://www.tobii.com/.
10. See http://www.eyetracking-glasses.com/ and http://www.smivision.com/.

Figure 13. Eye trackers; stationary system (left) and as eye glasses (right)

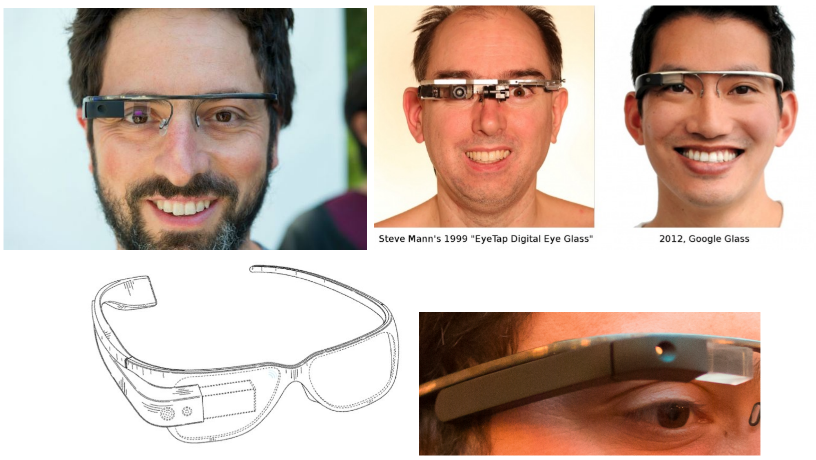


Figure 14. Examples of augmented reality glasses with camera

such glasses combine both presentation and gathering of data in one non-intrusive device.

The EPSON Moverio BT-100[11] is an Android-powered, see-through, wearable display, with the possibility of running apps on it. The technical features include side-by-side 3D technology, quarter-HD screen, interactive trackpad, smart navigation menus, and the Android platform. The Moverio allows users to view streaming video over WLAN networks, experience side-by-side 3D content, and enjoy downloaded digital content on a "floating" 80-inch perceived screen that is projected into the user's real environment. The EPSON Moverio BT-100 is shown in Figure 15.

11. See http://www.epson.com/cgi-bin/Store/jsp/Moverio/Home.do.

Figure 15. The Android-powered, see-through wearable display Moverio BT100

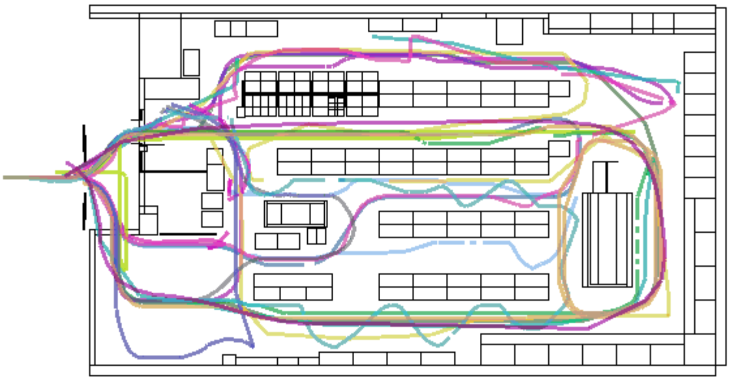


Figure 16. Behaviour assessment from movement tracking analysis

4.7 Behaviour Assessment from Movement Tracking

Movement tracking can be used for behaviour assessment. Leister et al. (2007b) show an example where movement tracking can be used for designing shops. In this example, a virtual shop was created using a game engine in order to create tracks of shoppers who were asked to perform a shopping task. These tracks, as shown in Figure 16, are then analysed with respect to optimising the users' path through a shop. Leister et al. also compare the user behaviour in the virtual store with user behaviour in real stores.

While outdoor positioning has been available over a decade for everybody to use, such as the GPS system, new technology for indoor-positioning[12] is emerging. A positioning system[13] determines the location of an object in a particular space. Real-time positioning systems can be used for location, tracking and location logging tasks in that space. Technologically, there are different technical solutions necessary to facilitate location capabilities in different spaces. Note that for some applications also the *orientation* and *direction* of an object could be relevant. Another relevant distinction is between *locating*, *tracking*, and *closeness to other objects*.

Global navigation satellite systems (GPS or GNSS) are generally not suitable to establish indoor locations, since microwaves are relatively weak, and will be attenuated and scattered by roofs, walls, and other objects.

Location and positioning systems can use different technologies which provide different accuracy. Which technology to choose depends on the application area, from coordinate measuring machines with an accuracy of tens of μm to GSM in the range of 100 m (Mautz, 2009). We also need to distinguish between technologies that require a setup and specific constraints in the movement pattern (e.g., systems based on interferometry), or whether arbitrary movements are allowed. An overview of indoor-location systems is given by Mautz (2012).

Commonly used technologies for indoor-tracking include:

---

12. See `http://en.wikipedia.org/wiki/Indoor_positioning_system` and `http://www.productivet.com/docs-2/Wireless_Comparison.pdf`.

13. The term *location system* would, in most cases, be more suitable.

(*a*) **Proximity to known points (choke point concepts):** This concept uses the proximity to devices where the position is known. The identification and proximity can be established using near-field communication (NFC) enabled devices and tags, e.g., using radio-frequency identification (RFID).

(*b*) **Grid concepts:** Placing many low-range receiver or sender devicess in a grid or other suitable pattern, such that it can be used for location determination purposes. By identifying the devices that are in range, the position can be estimated. Additionally, signal strength measurements can be used.

(*c*) **Long range sensor concepts:** These systems use a continuous physical measurement, e.g., length based on interferometry, along with the identification data in one combined signal.

(*d*) **Angle of arrival (AoA):** This technolgy uses arrays of highly directional sensors to determine the angle to an object. Triangulation to several transmitters can be used to calculate the position.

(*e*) **Time of arrival (ToA):** This technology uses the travel time of a signal to one or more targets. The GPS system is one example that uses trilateration; there are other possibilities, too.

(*f*) **Received signal strength indication:** The signal strength decreases with the distance from the emitter. Therefore, the signal strength can be used to measure the distance to one or several targets.

(*g*) **Inertial measurements:** Relative movements can be measured using inertial measurements, e.g., based on accellerometers (in smartphones), step counting (pedometers), etc.

(*h*) **Recognition of known landmarks:** The camera of a smartphone can be used to match the image of the environment to a model of the environment. In such a system the position and size of landmarks can be extracted and matched with the known position in a database. This approach requires image processing and image recognition, and search against a database of already registered objects. There are also approaches where specific patterns are projected onto the walls, and which are recognised by the image processing software in order to determine the location.

(*i*) **Sound-based:** This approach matches sound patterns received by a microphone and matches these against pre-recorded sound patterns for certain places. Possibly artificial sound patterns in a for humans inaudible frequency could be used (i.e., infrasound, ultrasound).

(*j*) **WLAN:** Locating by WLAN technology is a combinations of the technologies a, b, and f above. Also the MAC addresses of emitting WLAN base stations, together with a database of known positions of these base stations[14] can be used for location purposes.

14. Android phones can use this technology with the help of a database maintained by Google.

Several of the sensors needed to implement a location system are available in smartphones, and can be used to implement apps that allow to register location data. These location data can be used to study the behaviour of, e.g., museum visitors, and eventually get indications on the UX of these visitors.

4.8 Using Observations from Cameras

The *MIT Mood Meter* (Hernandez et al., 2012; Hoque et al., 2012) uses cameras to assess and display smiles related to the overall mood of passers by the installation. The installation gathers and aggregates facial affective information such as smiles, and replaces the faces on the camera image by corresponding icons that represent the mood. In an installation at the MIT, the numbers of smiles is supposed to represent a barometer of happiness. The installation is intended to raise awareness of how positive the effect of a smile can be, and, hence, represents an assessment method of UX.

Terzis et al. (2011) assess instant emotions based on facial expressions. Gorbunov et al. (2012) use the commercial software *FaceReader*[15] to extract seven time-dependent components of facial expressions and evaluate corresponding emotional stimuli using genetic programming.

Oliveira et al. (2011) are developing an emotional recognition engine named *iFelt*. The goal is to automatically recognise emotional states using psycho-physiological measures of the autonomous nervous system, including galvanic skin response, respiration, and heart rate. This tool is validated by the researchers using paper and pen observations and semi-structured interviews.

Another face recognition tool is *FaceReader* by the company *Noldus Information Technology*[16]. According to Terzis et al. (2010), *FaceReader* is capable of measuring emotions with a hit rate of over 87% during a self-assessment test, and that it could be successfully integrated into a computer-aided learning system for the purpose of affect recognition.

Yeasin et al. (2006) present a method of how to recognise six different facial expressions from visual data. These expressions are used to compute levels of interest. Yeasin et al. also give a literature overview of facial expression recognition. They mention that there are six emoticons—anger, fear, disgust, sadness, happiness, and surprise; lately, this list has been completed by including contempt and shame.

4.9 Quality of Context

Conflicting or incorrect results may occur through a large number of heterogeneous data sources, technical limitations in the sensors, aggregation, and defective sensor-generated data. Due to technical limitations in sensors and context reasoning algorithms, context information does not always represent accurately the reality, and *Quality of Context* (QoC) models have been proposed to quantify this inaccuracy (Neisse et al., 2008). Evaluation and verification of the QoC (Zimmer, 2006) is an important research challenge. The QoC

15. The FaceReader is commercial software developed by VicarVision and Noldus Information Technology.
16. See also `http://www.noldus.com/facereader/selected-publications`.

can be derived with similar estimation models as the QoE.

With the intrinsic heterogeneity and large number of sources of context, aggregating and reasoning on low-quality raw sensed data may result in conflicting and erroneous evaluation of situations. The evaluation and verification of QoC information will play a central role for applications depending on the consistency and correctness of provided context-aware data. This has been correctly emphasized by Arnrich et al. (2010) for eHealth applications, where the trade-off between a patient's comfort and signal quality is highlighted (e.g., an unobtrusive dry ECG electrodes embedded into clothes may offer a higher level of comfort to end-users than classical wet electrodes); however, the quality of the signal may be poor. Therefore, there is a strong need to apply automatic quality control mechanisms to raw sensed data, by running error compensation algorithms; e.g., error detection in ECG signal based on changes in the electrode-skin impedance (Ottenbacher et al., 2008) or on accelerometers (Gibbs and Asada, 2005). Indeed, it is useful to indicate the quality of context data as input for decision making, reinforcing the need for quality assessment and control.

References

Aaker, D. A., Stayman, D. M., and Hagerty, M. R. (1986). Warmth in advertising: Measurement, impact, and sequence effects. *Journal of Consumer Research*, 12(4):365–381. 18

Andrews, S., Bychkov, D., Crow, C., Kim, J., and Doornick, R. (2008). Detecting physiological responses in those exposed to real-life biochemical simulations. whitepaper. Available from: http://www.exmovere.com/pdf/Exmovere%20Chem-Bio%20Research-1.pdf. 22

Arnrich, B., Mayora, O., Bardram, J., and Tröster, G. (2010). Pervasive healthcare – paving the way for a pervasive, user-centered and preventive healthcare model. *Journal of Methods of Information in Medicine*, 49(1):67–73. 29

Batra, R. and Ray, M. L. (1986). Affective responses mediating acceptance of advertising. *Journal of Consumer Research*, 13(2):234–249. 18

Boucsein, W. (1992). *Electrodermal activity*. Plenum Press. 19

Brown, S. P., Homer, P. M., and Inman, J. J. (1998). A meta-analysis of relationships between ad-evoked feelings and advertising responses. *Journal of Consumer Research*, 35(1):114–126. 18

Brown, S. P. and Stayman, D. M. (1992). Antecedents and consequences of attitude toward the ad: A meta-analysis. *Journal of Consumer Research*, 19(1):34–51. 18

Bulling, A. and Gellersen, H. (2010). Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing*, 9(4):8–12. 24

Cacioppo, J. T., Berntson, G. G., Klein, D. J., and Hatfield, E. (1993). *The psychophysiology of emotion*, chapter 9, pages 119–142. R. Lewis, J. M. Haviland-Jones, and L. F. Barrett (Eds.): The handbook of emotions. Guilford, New York. 18

Cacioppo, J. T. and Gardner, W. L. (1999). Emotion. *Annu Rev Psychol.*, 50:191–214. 18

Chatting, D. J., Galpin, J. S., and Donath, J. S. (2006). Presence and portrayal: video for casual home dialogues. In *Proceedings of the 14th annual ACM international conference on Multimedia*, MULTIMEDIA '06, pages 395–401, New York, NY, USA. ACM. 18

Choi, M., Curtis, M. K., and McIntosh, H. (2009). Psychophysiological data acquisition in a virtual environment. In *SPIRE-EIT 09, Summer Program for Interdisciplinary Research and Education-Emerging Interface Technologies*. May 27, 2009 - August 1, 2009. Available from: https://www.hci.iastate.edu/REU09. 20

Cohen, J. B., Pham, M. T., and Andrade, E. B. (2008). *The Nature and Role of Affect in Consumer Behavior*, pages 297–348. Curtis P. Haugtvedt, Paul Herr, and Frank Kardes (Eds.): Handbook of Consumer Psychology. Tayler & Francis. 18, 20

Cutmore, T. R. and James, D. A. (2007). Sensors and sensor systems for psychophysiological monitoring: A review of current trends. *Journal of Psychophysiology*, 21(1):51–71. 18, 22

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011a). From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, pages 9–15, New York, NY, USA. ACM. 16

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., and Dixon, D. (2011b). Gamification. using game-design elements in non-gaming contexts. In *PART 2 – Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 2425–2428, New York, NY, USA. ACM. 16

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., and Dixon, D. (2011c). Gamification. using game-design elements in non-gaming contexts. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 2425–2428, New York, NY, USA. ACM. 16

Ebrahimi, T. and Sikora, T., editors (2003). *Visual Communications and Image Processing 2003*, volume 5150 of *Proceedings of SPIE*. SPIE. 35, 36

EBU project group B/VIM (2003). SAMVIQ - subjective assessment methodology for video quality. Report BPN 056 / B/VIM 053, European Broadcasting Untion (EBU). 10

Edell, J. A. and Burke, M. C. (1987). The power of feelings in understanding advertising effects. *Journal of Consumer Research*, 14(3):421–433. 18

Fabricatore, C., Nussbaum, M., and Rosas, R. (2002). Playability in action videogames: a qualitative design model. *Hum.-Comput. Interact.*, 17(4):311–368. 16

Ferreira, P. (2008). *Dealing with Stress: Studying experiences of a real-time biofeedback system*. PhD thesis, Department of Computer and Systems Sciences, Stockholm University / Royal Institute of Technology. 19, 22

Fukuda, R. (2011). Affective technology for older adults: does fun technology affect older adults and change their lives? In *Proceedings of the 6th international conference on Universal access in human-computer interaction: users diversity - Volume Part II*, UAHCI'11, pages 140–148, Berlin, Heidelberg. Springer-Verlag. 18

Gibbs, P. and Asada, H. H. (2005). Reducing motion artifact in wearable bio-sensors using MEMS accelerometers for active noise cancellation. *American Control Conference, 2005. Proceedings of the 2005*, 3:1581–1586. 29

González Sánchez, J. L., Padilla Zea, N., and Gutiérrez, F. L. (2009). From usability to playability: Introduction to player-centred video game development process. In *Proceedings of the 1st International Conference on Human Centered Design: Held as Part of HCI International 2009*, HCD 09, pages 65–74, Berlin, Heidelberg. Springer-Verlag. 16

Gorbunov, R., Barakova, E., and Rauterberg, M. (2012). Interpretation of time dependent facial expressions in terms of emotional stimuli. In *International Conference on Evolutionary Computation Theory and Applications*. Available from: http://www.amha.id.tue.nl/facial_expr_ecta2012.pdf. 28

Grimshaw, M., Nacke, L., and Lindley, C. A. (2008). Sound and immersion in the first-person shooter: Mixed measurement of the player's sonic experience. In *Proc. Audio Mostly*, Piteå. 20

Gross, J. J. and Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion*, 9(1):87–108. 18

Groven, A.-K., Haaland, K., Glott, R., Tannenberg, A., and Darbousset-Chong, X. (2012). Quality assessment of FOSS. In Leister, W. and Christophersen, N., editors, *INF5780 Compendium Autumn 2012: Open Source, Open Collaboration and Innovation*, volume DART/06/2012 of *NR Note*, chapter 5, pages 87–105. Norsk Regnesentral. 7

Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, 2:96–99. 18

Hernandez, J., Hoque, M. E., and Picard, R. W. (2012). Mood meter: large-scale and long-term smile monitoring system. In *ACM SIGGRAPH 2012 Emerging Technologies*, SIGGRAPH '12, pages 15:1–15:1, New York, NY, USA. ACM. 28

Hoene, C. (2005). Internet telephony over wireless links. Technical report, TU Berlin. 12

Holbrook, M. B. and Batra, R. (1987). Assessing the role of emotions as mediators of consumer responses to advertising. *Journal of Consumer Research*, 14(3):404–420. 18

Hoque, M. E., Hernandez, J., Drevo, W., and Picard, R. W. (2012). Mood meter: Counting smiles in the wild. In *Proc. UbiComp'12*, New York, NY, USA. ACM. 28

Horvath, K. and Lombard, M. (2010). Social and spatial presence: An application to optimize human-computer interaction. *PsychNology Journal*, 8(1):85–114. 18

Huynh-Thu, Q. and Ghanbari, M. (2008). Scope of validity of PSNR in image/video quality assessment. *Electronic Letters*, 44(23):800–801. 13

Ibáñez, J. (2011). Showing emotions through movement and symmetry. *Computers in Human Behavior*, 27(1):561–567. 19

IJsselsteijn, W., de Kort, Y., Poels, K., Jurgelionis, A., and Bellotti, F. (2007). Characterising and measuring user experiences in digital games. In *ACE Conference'07*, pages 1–4, Salzburg, Austria. 17

ITU-R (2004). Methodology for the subjective assessment of the quality of television pictures. Itu standardization document, ITU, Genève. 10

ITU Study Group 12 (2001). *Perceptual evaluation of speech quality (PESQ) – An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. ITU-T. 12

ITU-T (2003). *G.107: The E-model, a computational model for use in transmission planning*. ITU-T. 13

ITU-T (2005). *Single Ended Method for Objective Speech Quality Assessment in Narrow-Band Telephony Applications*. ITU-T. 12

Knee, M. (2006). The picture appraisal rating (PAR) - a single-ended picture quality measure for MPEG-2. White paper, Snell & Wilcox Limited. [Online]. Available: http://www.snellgroup.com/documents/white-papers/white-paper-picture-appraisal-rating.pdf, last accessed May 14, 2010. 13

Koumaras, H., Kourtis, A., Lin, C.-H., and Shieh, C.-K. (2009). End-to-end prediction model of video quality and decodable frame rate for MPEG broadcasting services. *International Journal On Advances in Networks and Services*, 1(1):19–29. [Online]. http://www.iariajournals.org/networks_and_services/ last accessed March 25, 2010. 13

Kozamernik, F., Steinmann, V., Sunna, P., and Wyckens, E. (2004). QSAMVIQ - a new EBU methodology for video quality evaluations in multimedia. *Proc. IBC 2004*. 10

Kozamernik, F., Sunna, P., Wyckens, E., and Pettersen, D. I. (2005). Subjective quality of internet video codecs. Ebu technical review, European Bradcasting Union (EBU). 10

Kristoffersen, T. (2012). Universal open content lifecycle management. Technical Report DART/18/2012, Norsk Regnesentral. 5

Lang, P. (1995). The emotion probe: Studies of motivation and attention. *American psychologist*, 50:372–372. 19

Larsen, J. T., Berntson, G. G., Poehlmann, K. M., Ito, T. A., and Cacioppo, J. T. (2008). *The psychophysiology of emotion*, chapter 11, pages 180–195. R. Lewis, J. M. Haviland-Jones, and L. F. Barrett (Eds.): The handbook of emotions. Guilford, New York, 3rd edition. 18

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P., and Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 719–728, New York, NY, USA. ACM. 7

Leister, W. (2012a). ICT-challenges in eHealth. NR Notat DART/15/2012, Norsk Regnesentral. confidential. 5, 6

Leister, W. (2012b). Research challenges for smart information systems. In *SMART 2012 (The First International Conference on Smart Systems, Devices and Technologies)*. International Academy, Research and Industry Association (IARIA). Available from: http://www.iaria.org/conferences2012/filesSMART12/Smart2012Keynote-WVL-V05HO.pdf. 5

Leister, W., Boudko, S., and Halbach, T. (2010). Estimation of subjective video quality as feedback to content providers. *Proc. ICSNC'10, Fifth International Conference on Systems and Networks Communication*, pages 266–271. 8, 11, 14, 15

Leister, W., Boudko, S., and Røssvoll, T. H. (2011). Adaptive video streaming through estimation of subjective video quality. *International Journal On Advances in Systems and Measurements*, 4(1&2):109–121. Available from: http://www.iariajournals.org/systems_and_measurements/index.html. 8, 11, 14, 15

Leister, W., Boudko, S., Steinmann, V., and Stoll, G. (2007a). Relationships between metrics of technical and perceived QoS for streamed video. Notat DART/07/07, Norsk Regnesentral. 8, 11

Leister, W. and Christophersen, N. D. (2012). Inf5780 compendium autumn 2012: Open source, open collaboration and innovation. Technical Report DART/06/2012, Norsk Regnesentral. Available from: http://publications.nr.no/1345627151/Compendium-INF5780H12.pdf. 5, 7

Leister, W. and Tjøstheim, I. (2012). Gamification, user experience, and sensors. In *CRM 2012*, Oslo. Den Norske Dataforening. Available from: http://publications.nr.no/1347549333/LeisterTjostheim-CRM2012-Handout.pdf. 5

Leister, W., Tjøstheim, I., and Lous, J. (2007b). Market research using a virtual test store on gaming technology. In Schulze, T., Preim, B., and Schumann, H., editors, *Simulation und Visualisierung 2007*, pages 399–411. ASIM - Arbeitsgemeinschaft Simulation, SCS Publishing House. 26

Liden, C. B., Wolowicz, M., Stivoric, J., Teller, A., Kasabach, C., Vishnubhatla, S., Pelletier, R., Farringdon, J., and Boehmke, S. (2010). Characterization and implications of the sensors incorporated into the sensewear armband for energy expenditure and activity detection. Technical report, BodyMedia, Inc. 24

Lombard, M. and Ditton, T. (1997). At the heart of it all: The concept of presence. *J. Computer-Mediated Communication*, 3(2):0. 17

Lombard, M., Ditton, T. B., and Weinstein, L. (2009). Measuring (tele)presence: The Temple Presence Inventory. In *Proc. Twelfth International Workshop on Presence*, Los Angeles, California. Available from: http://www.temple.edu/ispr/prev_conferences/proceedings/2009/Lombard_et_al.pdf. 18

Loosley, C. (2010). Apdex-G and Apdex-R composite sections 1-5. web pages. Available from: http://apdex.org/index.php/2010/09/apdex-g-and-apdex-r-composite-sections-1-5. 16

Malmstrom, E. J., Opton, E., and Lazarus, R. S. (1965). Heart rate measurement and the correlation of indices of arousal. *Psychosom Med*, 27(6):546–556. 20

Mautz, R. (2009). Overview of current indoor positioning systems. *Geodezija Ir Kartografija/Geodesy and Cartography*, 35(1):18–22. 26

Mautz, R. (2012). *Indoor Positioning Technologies*. habilitation thesis, ETH Zürich, Environmental and Geomatic Engineering, Institute of Geodesy and Photogrammetry. Available from: http://e-collection.library.ethz.ch/eserv/eth:5659/eth-5659-01.pdf. 26

Money, A. G. and Agius, H. (2009). Analysing user physiological responses for affective video summarisation. *Displays*, 30(2):59–70. 21

Mu, M. (2009). An interview with video quality experts. *SIGMultimedia Rec.*, 1(4):4–13. 13

Nacke, L., Stellmach, S., Sasse, D., and Lindley, C. A. (2009). Gameplay experience in a gaze interaction game. In Villanueva, A., Hansen, J. P., and Ersbøll, B. K., editors, *Proceedings of the 5th Conference on Communication by Gaze Interaction – COGAIN 2009: Gaze Interaction For Those Who Want It Most*, pages 49–54, Lyngby, Denmark. The COGAIN Association, The COGAIN Association. 17

Nacke, L. E. and Lindley, C. A. (2010). Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *CoRR*, abs/1004.0248. 16, 17, 20

Neisse, R., Wegdam, M., and van Sinderen, M. (2008). Trustworthiness and quality of context information. In *ICYCS*, pages 1925–1931. IEEE Computer Society. 28

NetPredict Inc. (2002-2004). *Assess the Ability of Your Network to Handle VoIP before You Commit*. NetPredict Inc. 13

Ni, P., Eichhorn, A., Griwodz, C., and Halvorsen, P. (2009). Fine-grained scalable streaming from coarse-grained videos. In *International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, pages 103–108. ACM. 13

Oliveira, E., Benovoy, M., Ribeiro, N., and Chambe, T. (2011). Towards emotional interaction: using movies to automatically learn users' emotional states. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part I*, INTERACT'11, pages 152–161, Berlin, Heidelberg. Springer-Verlag. 28

Ottenbacher, J., Kirst, M., Jatobá, L., Huflejt, M., Grossmann, U., and Stork, W. (2008). Reliable motion artifact detection for ecg monitoring systems with dry electrodes. *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society*, 2008:1695–1698. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19163005. 29

Paxson, V., Almes, G., Mahdavi, J., and Mathis, M. (1998). Framework for IP Performance Metrics. RFC 2330. [Online]. Available: http://www.ietf.org/rfc/rfc2330.txt, last accessed May 14, 2010. 12

Pezaros, D. P. (2005). *Network Traffic Measurement for the Next Generation Internet*. PhD thesis, Computing Dept. Lancaster University. 12

Picard, R. W. (2010). Emotion research by the people, for the people. *Emotion Review*. 23

Pinson, M. H. and Wolf, S. (2003a). Comparing subjective video quality testing methodologies. In Ebrahimi and Sikora (2003), pages 573–582. 10

Pinson, M. H. and Wolf, S. (2003b). An objective method for combining multiple subjective data sets. In Ebrahimi and Sikora (2003), pages 583–592. 10

Poh, M.-Z., Swenson, N. C., and Picard, R. W. (2010). A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity. *Biomedical Engineering, IEEE Transactions on*, 57(5):1243–1252. 23

Rapson, R. L., Hatfield, E., and Cacioppo, J. T. (1993). *Emotional Contagion*. Cambridge University Press. 18

Reynolds, C. J. (2005). *Adversarial Uses of Affective Computing and Ethical Implications*. PhD thesis, Massachusetts Institute of Technology. 23

Roto, V., Law, E., Vermeeren, A., Hoonhout, J., et al. (2011). User experience white paper. Technical report, Result from Dagstuhl Seminar on Demarcating User Experience, September 15-18, 2010. Available from: www.allaboutux.org/uxwhitepaper. 7

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178. 19

Sanches, P. A. N. (2008). Supporting self-reflection in everyday life: An exploratory review of physiological input methods for the affective health system. Master's thesis, Department of Computer and Systems Sciences, Stockholm University / Royal Institute of Technology. 19, 20

Schachter, S. and Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5):379–399. 21

Sevcik, P. (2005). Defining the application performance index. *Business Communications Review*, pages 8–10. 15

Smotlacha, V. (2001). QoS oriented measurement in IP networks. Report 17/2001, CESNET. [Online]. Available: http://www.cesnet.cz/doc/techzpravy/2001/17/qosmeasure.pdf, last accessed May 14, 2010. 12

Steinmann, V., Vogl, A., and Stoll, G. (2006). *Subjective assessment of video quality depending on encoding parameters, Contribution of IRT to MOVIS WP2 Phase 1*. Institut für Rundfunktechnik. 11

Stoll, G., Vogl, A., and Steinmann, V. (2007). *Subjective Assessment of Video Quality depending on networking parameters*. Institut für Rundfunktechnik. 11

Tamás, G., Clements-Croome, D., and Wu, S. (2007). The use of wireless data communication and body sensing devices to evaluate occupants' comfort in buildings. In *Proc. CLIMA 2007 - Wellbeing Indoors*, Helsinki. 24

Tektronix (2000). A guide to maintaining video quality of service for digital television programs. Technical report, Tektronix Inc. 9

Terzis, V., Moridis, C. N., and Economides, A. A. (2010). Measuring instant emotions during a self-assessment test: the use of FaceReader. In *Proceedings of the 7th International*

Conference on Methods and Techniques in Behavioral Research, MB '10, pages 18:1–18:4, New York, NY, USA. ACM. 28

Terzis, V., Moridis, C. N., and Economides, A. A. (2011). Measuring instant emotions based on facial expressions during computer-based assessment. *Personal and Ubiquitous Computing*, pages 1–10. 28

Tsui, K. M., Desai, M., and Yanco, H. A. (2012). Towards measuring the quality of interaction: communication through telepresence robots. In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*, PerMIS '12, pages 101–108, New York, NY, USA. ACM. 18

Vittorias, I., Petrantonakis, P., Bolis, D., Tsiligkyri, A., Kosmidou, V., and Hadjileontiadis, L. J. (2008). Noesis: An enhanced educational environment for kids with autism spectrum disorders. In *ICALT*, pages 1019–1020. IEEE. 22

Wang, Y. J. and Minor, M. S. (2008). Validity, reliability, and applicability of psychophysiological techniques in marketing research. *Psychology & Marketing*, 25(2):197–232. 18

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612. 13

Wassermann, A. I., Pal, M., and Chan, C. (2005). BRR whitepaper 2005 RFC 1. Technical report, Business Readiness Rating Project, Carnegie Mellon West. currently not available, November 18, 2010. Available from: http://www.openbrr.org/wiki/images/d/da/BRR_whitepaper_2005RFC1.pdf. 7

Wilhelm, F. H. and Grossmann, P. (2010). Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, 84:552–569. 19, 20

Wolf, S. and Pinson, M. (2002). Video quality measurement techniques. Ntia report 02-392, National Telecommunications and Information Administration, US Dept of Commerce. 13

Yeasin, M., Bullot, B., and Sharma, R. (2006). Recognition of facial expressions and measurement of levels of interest from video. *Multimedia, IEEE Transactions on*, 8(3):500–508. 28

Zimmer, T. (2006). QoC: Quality of Context – improving the performance of context-aware applications. In *Advances in Pervasive Computing. Adjunct Proceedings of Pervasive 2006*. APC. 28