Context-Aware QoE Modelling, Measurement, and Prediction in Mobile Computing Systems

Karan Mitra, Arkady Zaslavsky, and Christer Åhlund

Abstract—Quality of Experience (QoE) as an aggregate of Quality of Service (QoS) and human user-related metrics will be the key success factor for current and future mobile computing systems. QoE measurement and prediction are complex tasks as they may involve a large parameter space such as location, delay, jitter, packet loss, and user satisfaction just to name a few. These tasks necessitate the development of practical context-aware QoE models that efficiently determine relationships between user context and QoE parameters. In this paper, we propose, develop, and validate a novel decision-theoretic approach called CaQoEM for QoE modelling, measurement, and prediction. We address the challenge of QoE measurement and prediction where each QoE parameter can be measured on a different scale and may involve different units of measurement. CaQoEM is context-aware and uses Bayesian networks and utility theory to measure and predict users' QoE under uncertainty. We validate CaQoEM using extensive experimentation, user studies and simulations. The results soundly demonstrate that CaQoEM correctly measures range-defined QoE using a bipolar scale. For QoE prediction, an overall accuracy of 98.93% was achieved using 10-fold cross validation in multiple diverse network conditions such as vertical handoffs, wireless signal fading and wireless network congestion.

Index Terms—Bayesian networks, context-awareness, decision theory, quality of experience, simulations, prototyping, user tests

1 INTRODUCTION

I N mobile and pervasive computing environments, understanding and measuring users' quality of experience (QoE) is an important and challenging task [8], [17], [22], [41]. There are several stakeholders who are interested in understanding what users think and perceive about the services being provided to them in terms of new products and applications [8]. For example, telecommunications operators want to understand how to minimize network churn by providing better services to the users. On the other hand, network engineers require the knowledge about underlying network conditions affecting users' QoE for user-centric network optimization.

Users' QoE about a particular technology, network service or an application depends on their expectations, on their cognitive, psychological and behavioural states and network parameters which vary with time and under different context and situations [8], [12], [17], [22]. For example, Jumisko-Pyykkö and Hannuksela [21] shows that users' QoE may differ in laboratory and real-life user environments. It is widely assumed [17], [21], [29], [41] that by maximizing network QoS (e.g., increasing network bandwidth and/or increasing wireless signal strength) or by

Manuscript received 19 Mar. 2013; revised 24 Aug. 2013; accepted 30 Oct. 2013. Date of publication 5 Dec. 2013; date of current version 30 Mar. 2015. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier 10.1109/TMC.2013.155

reducing the cost of services, users will be satisfied with the services provided to them. It is argued by [17], [21], [22], [41] that QoS provided by the operators may not correlate well with users' QoE. For example, consider a statement posted on the Apple forum:¹

Example 1. "I am having the same issues as everyone else...Phone shows 5 bars on "4G." Makes calls and texts just fine but no imessage or internet (safari as well as any other apps that require connectivity). Right now the two things I have noticed are that I'm more likely to have it work late at night (11pm-2am) and more likely to have it work when I'm outdoors. Most of my problems are experienced while at work (9am-6pm) and when I leave work it tends to work for a while (lunch break, errands) but not always. I did try turning off all the 'System Services' under 'Location services' and have not noticed any difference in my phone's behavior."

This example shows that positive user experience may not be guaranteed even with 4G networks. In this paper, we argue that QoE is a broader construct which encompasses QoS parameters along with user related factors such as their behavioural, cognitive and psychological states along with the context in which these products and services are provided to them. These factors determine their overall QoE related to a particular application or service. In mobile computing environments, users' behaviour is dynamic as they use applications or services in different scenarios. Thus, for QoE measurement and prediction, it is important to consider parameters related to users, their device(s) and environment along with network QoS. For example, information such as user's moods, screen size, social context,

K. Mitra is with Monash University, Melbourne, VIC 3145, Australia, and also with Luleå University of Technology, Luleå, Norrbotten 97187, Sweden. E-mail: karan.mitra@ltu.se.

A. Zaslavsky is with CSIRO, Canberra, ACT 2601, Australia, and also with Luleå University of Technology, Luleå, Norrbotten 97187, Sweden. E-mail: arkady.zaslavsky@csiro.au.

C. Åhlund is with Luleå University of Technology, Luleå, Norrbotten 97187, Sweden. E-mail: christer.ahlund@ltu.se.

^{1.} https://discussions.apple.com/thread/3437795?start=180& tsta rt=0. Retrieved 11/07/12.

^{1536-1233 © 2013} IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

activities and packet loss. Thus, based on broad features of QoE, we define it as:

Definition 1. "Quality of experience (QoE) is a metric that depends on the underlying QoS along with a person's preferences towards a particular object or service where his/her preferences are defined by his/her personal attributes related to expectations, experiences, behaviour, cognitive abilities, object's attributes and the environment surrounding that person".

Context is any information that assists in determining a situation(s) related to a user, network or device [13]. For example, from GPS coordinates, user-related situation can be inferred as: "user is at work". From delay of 50 ms and packet losses of 0%, network-related situation can be determined as "network is not congested". We consider context as any information that assists in determining users' QoE. Context can be static and dynamic. Static context does not change often, while dynamic context changes over a period of time and is difficult to predict. Static context may include user's application preferences, their security requirements and cost. In real-life environments, context can be highly dynamic and stochastic i.e., it can change in a very short period of time and is uncertain; it can be imperfect; it can exhibit a range of temporal characteristics; it can have several alternative representations; it can be interrelated; it can be distributed; and it may not be available at a particular time [7]. The timely collection and processing of context may be crucial as it may loose its accuracy. Dynamic context may include user location, velocity, network load, battery power, memory/CPU utilization, presence and signal to noise ratio. We see QoE as a function of context attributes (represented as a vector). This function is written as:

$$QoE = f(context \, vector) \tag{1}$$

Table 1 shows several context attributes that may assist in QoE measurement and prediction. Studying these attributes together and understanding their effects on each other is a highly complex task [8], [37], [41] as they may contain hidden inter-dependencies and non-linear relationships. There is a need to develop QoE measurement and prediction models that consider several attributes (both subjective and objective) under different situations related to a user, device, application and network to measure and predict QoE [8], [41], possibly on a single scale [8]. Once developed, these QoE models may benefit stakeholders such as network operators to minimize network churn and thereby maximizing their revenues. These models can also be used to provide users with personalized services and applications on their mobile devices.

Recently, several researchers [8], [12], [18], [21], [22], [24], [35], [37], [38], [41] considered the challenges associated with QoE measurement and prediction. However, they do not deal with the challenge of building a practical context-aware QoE model that models several context attributes and QoE parameters for efficient QoE measurement and prediction. Also, the challenge of QoE measurement on a single scale by combining various QoE parameters is not addressed. Current approaches do not deal with the

TABLE 1

Context Attributes Related to User, Device, Application and Network for Modelling, Measuring, and Predicting QoE

Туре	Context parameters
User and user	location, temperature, heart rate,
environment	eye movement, amount of sweat,
	social context, people nearby,
	light, background noise, age,
	gender
Tool/device/	screen size, design layout,
object	resolution, general intuitiveness,
	buttons placement, input/output
	methods, appeal, usability
Application	type, requirements
Network	type, bandwidth, delay, jitter,
	packet loss, RTT, packet loss burst
	size, protocols used, received
	signal strength, congestion levels

problem of imprecision in QoE measurement due to inherent uncertainty caused by scarce data or other uncontrollable factors prevalent in mobile computing environments. We assert that these challenges are not mutually exclusive and should be considered together to realize efficient methods for QoE measurement and prediction.

1.1 Our Contributions

In this paper, we propose, develop and validate a novel Context-aware approach for Quality of Experience modelling, Measurement and prediction (CaQoEM). In particular, our contributions are two-fold.

- 1) We propose, develop and validate a context-aware method for QoE modelling based on context spaces model (CSM) [31] and Bayesian networks [19], [36]. Our method considers several context attributes and QoE parameters. It provides a simple and efficient way to determine relationships between these attributes as well as parameters to measure and predict users' QoE.
- 2) We propose, develop and validate a context-aware, decision-theoretic method for QoE measurement and prediction under uncertainty on a single scale. We validate CaQoEM using a number of case studies. We developed a prototype and performed a number of subjective tests. Finally, numerical analysis using extensive simulations in OPNET [3] was performed. We show that CaQoEM predicts users' QoE with high accuracy of 98.93% under multiple diverse network conditions prevalent in mobile computing systems.

CaQoEM is a novel, extensible and a unifying approach to consider the challenge of context-aware QoE modelling, measurement and prediction. This paper is organized as follows. Section 2 presents the related work. Section 3 presents our proposed approach. Section 4 presents results analysis and finally, Section 5 presents the conclusion and future work. Quality of experience (QoE) measurement techniques can be classified into subjective and objective [8], [12], [38], [41]. Subjective measurements mainly consider direct data collection from end users [8], [12]. For example, asking a user to rate the quality of multimedia stream for applications such as VoIP [26] and IPTV [21]. However, there are several problems that arise while considering subjective tests. These tests can be too expensive and time consuming to be performed in a university or even in a small industry environment [8], [26], [38]. Hence, these tests are mainly limited to major telecommunication operators. The standardization bodies such as the ITU-T in its ITU-T P.800 recommendation [30] have defined methods to measure users' subjective quality based on a score called the mean opinion score (MOS).

Several researchers [8], [18], [25] noted problems while adhering to the ITU-T P.800 recommendation for conducting subjective tests. For instance, the major problem while considering MOS as a QoE metric is that it considers arithmetic computation involving user ratings. We assert that arithmetic operations such as computing mean and standard deviation cannot be applied on subjective ratings as these ratings are categorical in nature. The human test subjects ranks the alternatives (e.g., "excellent" and "fair") on the categorical/ordinal scale where the distance between these alternatives cannot be known [8], [16], [18], [25]. Thus, the methods involving MOS as a metric will also be inappropriate [16], [25], [27]. Mu et al. [27] show that in case of subjective tests, normality of collected data (user ratings) cannot be verified. Further, due to the subjective nature of users' ratings, parametric statistical models cannot be applied for QoE measurement and prediction. Hence, techniques involving ordinary regression analysis will also be invalid. The authors point out that the conditions for valid statistical tests are rarely verified and documented by computer science researchers. Nonetheless, MOS is the most widely used method to assess subjective ratings in both industry and academia.

Takahashi et al. [38] argued for developing objective methods to predict subjective quality for multimedia applications. Objective tests can be considered as a cheaper means for quality assessment and prediction as test subjects (users) are not required for QoE measurement. Telecommunication systems research in the past has used objective methods to assess VoIP call quality without user intervention. However, these methods, for example, the ITU-T E-Model [14] are hard to model and develop due to large parameter space which causes less accuracy in quality assessment. Further, most of the objective methods were developed under strict laboratory environments and with a limited number of objective and subjective parameters such as delay, jitter, packet loss, and bandwidth. Authors in [26] show that evaluating and extending the objective methods such as the ITU-T E-Model [14] is challenging as it involves several dependent and independent parameters. In case any changes are made to the current model, new subjective tests will have to be performed and new mathematical models will have to be developed [8].

To alleviate such problems, Menkovski *et al.* [24] and Rubino, Tirilly and Varela [35] proposed artificial intelligence (AI)-based methods for QoE prediction. However, in case of scarce and sparse data, methods such as [24] may fail as they simply consider classification accuracy as a metric for evaluation. For example, consider a case where six users out of ten gave '5' as a rating and the remaining four users gave '4' as a rating to VoIP call quality. In this case, [24] will select QoE as '5' with probability of 0.60. This is incorrect as it ignore ratings of other four users.

These methods may not give accurate QoE estimations in real-life environments as users' OoE is also dependent on other context attributes related to their expectations [22], [25], [37] and their behavioural and psychological states [41]. Moor et al. [12] argues for a multi-disciplinary approach for QoE measurement and prediction. Their approach involves living-labs testing where QoE measurements are done through the use of software tools in realistic user test environments. Jumisko et al. [21] conducted a study from mobile IPTV perspective in living-labs setting. Their results show much lower error rates in users' QoE ratings in case of the realistic test environments as compared to the laboratory settings. However, the problem with QoE measurement and prediction in real-life environments is that most of the parameters are uncontrollable. Further, users' behaviour and network conditions can be dynamic and uncertain.

Brooks and Hestnes [8] stress the need to consider a combination of a number subjective and objective parameters to determine QoE on a single scale. However, we gather that there is a death of methods, both subjective and objective, that consider the problem of QoE measurement and prediction on a single scale by considering several QoE parameters. From the state-of-the-art, we identified the following challenges for QoE measurement and prediction:

- Context-aware QoE modelling: QoE measurement 1) and prediction may involve a large parameter space comprising of several context attributes and QoE parameters as shown in Fig. 13. There can be N context attributes affecting M QoE parameters. Further, M QoE parameters can affect each other. Thus, selecting relevant context attributes and QoE parameters and finding relationships between them can be challenging. These relationships are usually non-linear and hard to quantify. This necessitates the development of novel QoE modelling techniques that model several context attributes and QoE parameters efficiently. Contrary to QoE models presented in [11], [22], [23], the new QoE models should not only be conceptual, but should also transcend to solving the challenges associated with QoE measurement and prediction. For example, rather than simply classifying and representing the parameters, QoE models should directly be used for QoE measurement and prediction.
- 2) **Context-aware QoE measurement and prediction:** The challenge of QoE measurement and prediction involving multiple context attributes and QoE parameters is not well addressed. Each QoE parameter can be measured on a different scale and



Fig. 1. Our high-level approach for QoE measurement and dissemination to relevant stakeholders.

may involve different units of measurement [8], [25]. These scales can be qualitative or quantitative. For example, QoE parameter "user satisfaction" is measured using an ordinal (qualitative) scale involving ratings 1 to 5 or 1 to 9. Thus, determining QoE based on different types of scale and/or different units of measurement remains a challenging task. Further, current methods do not explicitly deal with the problem of imprecision in QoE measurement and prediction due to uncertainty caused by scarce and sparse data or other uncontrollable factors prevalent in both laboratory and real-life environments.

In the next section, we address the aforementioned challenges and present our approach, CaQoEM for QoE modelling, measurement and prediction.

3 CAQOEM: CONTEXT-AWARE QUALITY OF EXPERIENCE MEASUREMENT AND PREDICTION IN MOBILE COMPUTING SCENARIOS

This section presents our approach for context-aware QoE modelling, measurement and prediction. Our high-level approach is presented in Fig. 1. Firstly, context attribute values pertaining to bandwidth, user satisfaction, screen size and people nearby are collected from the network, user, device and the user's surrounding environment. Secondly, context modelling is performed on these attributes using techniques such as Bayesian networks (BNs) [19], [36]. Lastly, QoE is inferred and measurement using a decision-theoretic approach comprising of BNs, utility theory [39] and a bipolar scale. The measured QoE is presented to the stakeholders for further analysis and use.

3.1 Context Modelling for Quality of Experience Inference Using Bayesian Networks

We extend the context spaces model (CSM) [31] and use Bayesian Networks (BNs) for modelling context information. The motivation behind CSM is to represent and reason about real-life situations related to users or applications in a naturalistic manner. It provides a method to deal with conflicting information and to reason about situations in uncertain conditions. The concept is similar to BNs where



Fig. 2. Context spaces model for QoE modelling.

the information is collected and is used to infer certain hypotheses probabilistically. In CSM, as shown in Fig. 2, situations are determined from raw context attribute values collected from sensors such as GPS and QoS probes. The raw context values are first used to infer context states which are *fused* to determine an overall situation. There are three important concepts in CSM:

Definition 2. A context attribute is the data element at time t that is used to infer a situation(s). It is represented as a_i^t . Primary context attributes are related to physical and virtual sensors such as GPS to collect location coordinates and signal strength measurement by the operating system. Secondary context attribute requires assistance of primary context attribute(s) to determine its value. For example, user's mood context attribute value can be determined using primary context attributes such as heart rate and location (see Fig. 13).

Research areas such as Affective Computing and Human Computer Interaction deal with the inference of users' mood. For example, [32], [34] show that different types of moods and emotions can be inferred using both physiological (blood pressure and heart rate) as well as other context attributes such as location, activity and voice. These primary context attributes can also be incorporated within our research to determine mood as a secondary context attribute.

- **Definition 3.** A context state refers to the current state of a user, application, network or device being modelled at time t based on the context attributes. It is represented as $S_i^t = (a_1^t, \ldots, a_n^t)$ where $n \in N$ represents n^{th} context attribute. For example, consider a VoIP application experiencing the delay (a_i^t) of 20 ms and 0% packet loss (a_j^t) . The resultant mean opinion score (MOS) calculated using the E-Model [14] would be above 4 (out of 5). Thus, context state, user satisfaction (S_i^t) is "very good". If the application is used by a user at his/her home, another context attribute, location (a_k^t) can be added to the context state to infer a different situation².
- **Definition 4.** A situation space (R) reflects a real-life situation. It is a collection of regions of attribute values corresponding to a predefined situation. It is denoted as a set $R^t = \{C_1^R, \ldots, C_n^R\}$ and $a_i^t \in C_i^R$ where C_i^R represents a set of

2. For the sake of clarity, we use the notations such a a_{delay}^t and $a_{location}^t$ to represent context attributes instead of using indices such as i, j and k. Similarly, a context state "user satisfaction" and "technology acceptance" can be written as S_{US}^t and S_{TA}^t instead of S_i^t and S_j^t .



Fig. 3. Bayesian network representation of context spaces model.

context attributes taking numerical or non-numerical values in an acceptable range. For example, context attribute, delay (a_{delay}^t) , can take a value in the range of 0 to 150 for context state "user satisfaction is excellent". On the other hand, a_{delay}^t takes a value within the range 450 to ∞ for context state "user satisfaction is poor". Situations can be defined for users such as "users' overall QoE is fair" or for the network such as "a network is congested".

To design an experiment for QoE measurement, stakeholders can think about context attributes such as $a_{location}^{t}$, a_{delay}^{t} and $a_{packetloss}^{t}$ that affect context states such as S_{US}^{t} and S_{TA}^{t} and possibly, the overall situation R_{QoE}^{t} . To reason about situations, several machine learning techniques such as Bayesian networks (BNs), fuzzy logic theory (FLT) and Decision Trees (DTs) can be applied. For QoE measurement, we need to deal with users' subjective ratings reflecting their behaviour and expectations. Thus, it is imperative to choose techniques that provide naturalistic way of thinking, representing and reasoning about situations such that it is easier for the stakeholders to model, measure and predict users' QoE.

In this paper, we consider BNs as they can learn efficiently from user ratings and can deal with uncertainty caused by scarce and sparse data. Further, BNs can handle both numerical and non numerical data. Unlike DTs [24] and random neural networks (RNNs) [35], BNs can incorporate experts knowledge in the form of domain knowledge. Compared to FLT, BNs do not require formulation of explicit rules to reason about situations. BNs can be extended to dynamic Bayesian networks (DBNs) to reason about situations over time. Last but not the least, BNs can be used with utility theory to make decisions under uncertainty [19], [36]. We now show that BNs can be used to model human behaviour effectively in uncertain mobile computing scenarios and can be used for QoE measurement and prediction. A BN is defined as follows:

Definition 5. A Bayesian network (BN) is a directed acyclic graph (DAG) where, random variables form the nodes of a network. The directed links between nodes form the causal relationships. The direction of a link from X to Y means that X is the parent of Y. Any entry in the network can be calculated using the joint probability distribution (JPD) denoted as [36]:

$$P(x_i, \dots, x_n) = \prod_{i=1}^n P(x_i | Parents(X_i))$$
(2)

where, parents nodes X_i , is the parent of node x_i . Fig. 3 gives the BN representation of the CSM [31] shown previously in Fig. 2. As can be seen in Fig. 3, there is a strong



Fig. 4. Bayesian network for QoE inference and prediction.

resemblance between CSM and our BN i.e., the concepts of CSM such as context attributes, context states and situations are represented as a DAG. This model presents a rich description of how context attributes are related to context states which are subsequently related to the situations. This can be seen as *causal model* where the arcs from context attributes to context states represent their effect on the context states. Also, the arcs from context states to the situations suggest that they cause a particular situation to occur. It can be appreciated that *N* number of context states and attributes can be added to this BN. This is beneficial when multiple QoE related context states need to be inferred together.

Fig. 4 shows a simple BN for QoE inference and prediction. At the lowest level, values for context attributes such as bandwidth $(a_{bandwidth}^t)$ and location $(a_{location}^t)$ are collected. At the intermediate level, context states such as technology acceptance (S_{TA}^t) and user satisfaction (S_{US}^t) are inferred. These states are represented as "hidden nodes" as we do not observe users' intention directly [42]. The topmost state is the situation (R_{OoE}^t) which is also a hidden node. It is used to determine the overall QoE situation of users and is dependent on context states below it. There are several other possible ways of constructing a BN such as a naive Bayes model or a more complex model built on causality [19], [36]. It is worth mentioning here that causal relations between the nodes of a BN may sometimes be difficult to be defined by stakeholders. One test to check whether two nodes (say, node A and node B) are related is: if someone fixes the state of a node A, and if that does not change the belief of node B then, A is not a cause of *B* [19]. For example, let us consider a context attribute a_{delay}^t and a context state S_{US}^t . If a change occurs in the value of a_{delay}^{t} and it causes some changes in beliefs of S_{US}^{t} , it means both nodes are related otherwise they are not. For the sake of brevity, we do not discuss creating various BNs in this paper. Interested readers can refer to [19], [36] for thorough discussion on this topic.

Each discrete node in a BN has a conditional probability table (CPT). In case of continuous nodes, conditional probability distribution (CPD) is defined. Usually Gaussian distribution is considered but any other distribution can be considered [36]. The CPTs can be learnt from the data or can be defined subjectively by the field/domain experts. In some cases, setting of CPTs may be quite challenging even if robust statistical models are present. In such cases, the maximum entropy methods can be used with BNs to assist experts in fitting the CPTs in a BN more precisely [40]. Stakeholders can also use structural learning algorithms



Fig. 5. Bayesian network (BN_1) and utility nodes (shown as diamond nodes) to determine context states such as "user satisfaction" and "technology acceptance" and the situation state (QoE).

such as structural expectation maximization and Markov chain Monte Carlo [28], [36] to create a BN which will reflect the relationships between context attributes, context states and the situation. For now, let us simply assume the model shown in Fig. 4. We will show in Section 4 that a simple model can be used to infer and predict QoE situations with accuracy greater than 95%.

As discussed previously, there can be N context attributes belonging to M QoE states. Using Eq. 2, context attributes such as $a_{bandwidth}^t$ and $a_{location}^t$ can be used to infer context states S_{US}^t and S_{TA}^t . Similarly, M context states can be inferred simultaneously without the need for a precise mathematical model. Once we have inferred all context states, these can be fused together to determine the overall QoE. By fusing, we mean to determine the probability distribution of QoE states and then computing a numerical value that determines the overall QoE. As our aim is to determine QoE on a single scale based on multiple QoE states, such as S_{US}^t and S_{TA}^t , it is important to map these states onto an objective function based on the requirements of stakeholders. Once we have determined the hypotheses of multiple QoE states, we need to determine what they actually mean together. For example, if "user satisfaction" in terms of call quality is "good" while the user is "not willing to accept the technology", what does this statement mean? One situation can be that "user is not happy with the overall QoE being offered to him/her". This arises the challenging task of correctly determining an overall QoE situation. Even with this simple example, we can realize the complexity of making a correct decision. We gather that there is inherent uncertainty associated with: a.) determining the hypotheses of each QoE state. For this we use BNs; and b.) in making an overall decision regarding QoE. For this, we use decision-theory [36], [39]. In the next section, we discuss our decision-theoretic approach for QoE measurement and prediction.

3.2 Bayesian Networks and Utility Theory for Quality of Experience Measurement

This section presents CaQoEM for QoE measurement and prediction. Fig. 5. shows our overall approach. At the lowest level, values for context attributes such as $a_{packetloss}^{t}$ and $a_{location}^{t}$ are collected. At the intermediate level, context

states such as S_{US}^t and S_{TA}^t are probabilistically inferred. As we aim to find a single numerical value to measure the overall QoE, utility value (utility nodes are shown as diamond nodes) for each context state is computed. These utility values are added as a linear weighted sum to determine a single scalar value. This scalar value is mapped to an interval scale to determine the overall QoE situation (R_{OoE}^t).

In CaQoEM, each QoE parameter is represented as a context state in a BN (see Figs. 4 and 5). Firstly, using a BN, stakeholders can infer the probability of hypothesis pertaining to each alternative. For example, $P(S_{IIS}^t =$ "excellent") = 0.90 and $P(S_{US}^t = "very good") = 0.10$. Then using utility theory, a single scalar value (expected utility (EU)) is determined. We assert that QoE measurement by simply considering the hypothesis with highest probability may be incorrect especially in case of small datasets. For example, consider a case where $P(S_{US}^{t} =$ "excellent") = 0.51 and $P(S_{US}^t = "very good") = 0.49$. The BN will classify S_{US}^t as "excellent" which is incorrect as we have ignored users who have selected "very good" as the outcome. This necessitates the need to consider all possible hypotheses. In this section, we will discuss how we can deal with this problem by incorporating utility theory [39] along with BNs to correctly infer and measure the overall OoE.

The idea is to derive a scalar value for each context state such as S_{SA}^t and S_{TA}^t . Later, these context states can be fused together mathematically to determine the overall QoE situation (R_{QoE}^t). The idea of deriving scalar values for each context state solves two problems regarding QoE measurement. Firstly, it makes it possible to perform mathematical operations such as taking a mean or standard deviation. These mathematical operations cannot be applied directly in most of the QoE measurement techniques where rank ordered scales such as the Likert and absolute category scales (ACR) are considered [16], [18], [25], [27]. Lastly, using these scalar values, we have a chance to deal with imprecision and uncertainty as we get a finer control to map these values according to any scale we like.

It is important to understand that when human subjects give a rating regarding their QoE, they usually prefer one outcome over another. For example, on a 5-point Likert scale, users give a rating of '5' for "excellent" and '1' for "poor". Thus, users prefer one outcome over another and try to give maximum value to best alternative on the scale. Thus, we can directly use the principle of maximum expected utility (MEU) [36], [39] to derive the scalar value for each context state. The MEU principle states that a rational agent (users) should choose an action (give rating) that maximizes the agent's expected utility (of each context state such as S_{US}^t) [39]. To measure the overall QoE on a single scale based on several QoE states, we define a global utility function (GU(QoE)) which can encompass in a generic way all context attributes and states.

Based on decision theory [19], [36], we define the alternatives $(a \in A)$, hypotheses $(h \in H)$ and evidence variables (e). $P(\bullet)$ represents the belief of an agent in a hypothesis and $U(\bullet)$ encodes the preference on the numerical scale. For S_i^t , h and each alternative a, we have a U(h, P(h)) describing

926



Fig. 6. Interval scale for QoE measurement.

the utility of *h* if it belongs to context state, S_i^t . Based on this, the expected utility for an individual QoE state (S_i^t) is computed as:

$$EU\left(S_{i}^{t}\right) = \sum_{h \in H} U(h)P(h) \tag{3}$$

Once the utility for each QoE state is computed using Eq. 3, we need to map this utility value on a scale to select an alternative, a. This alternative is the QoE for that state. Since we are dealing with human judgements in the form of subjective ratings, we need to consider a proper scale of measurement. Typical scales for this purpose include the ordinal or interval scale [8], [16]. Dealing with an interval scale in case of human judgements can be difficult therefore, ordinal scale is mostly used [16]. For example, ITU-T P.800 recommendation [30] considers the ordinal scale for conducting subjective tests. This scale ranks the alternatives on the scale therefore, arithmetic operations such as computing average and standard deviation cannot be applied [8], [25], [27], as discussed previously. To solve this problem, we propose to use a bipolar scale to convert qualitative attributes (QoE) into an interval scale [16]. This conversion is performed by using linear scale transformation (discussed below). For it's proper calibration, we propose to use the end-points method where the maximum value ('5' as "excellent") is assigned as '1' and the minimum value ('1' as "poor") is assigned as '0'. The midpoint, '0.50' is used for calibration since it is the breakpoint between favourable (i.e., better than average values) and unfavourable values (i.e., below average values).

The challenge here is to map the so called qualitative alternatives such as "excellent", "very good", "good", "fair" and "poor" to an interval scale. As these alternatives are qualitative, we need to carefully assign them to the intervals on the scale. For example, "excellent" can be assigned to values greater than 0.8751 and "poor" can be assigned values in the range: [0:0.1250]. Similarly, "fair" can be assigned the range: [0.1251:0.3750], "good" can be assigned the range: [0.6251:0.8750] as shown in Fig. 6. Once these alternatives are mapped onto the intervals, it is easy to perform mathematical operations [8], [16], [25] such as addition and multiplication to determine the context states and QoE situation.

To compute the expected utility for each context state, we assign a utility value to each hypothesis of the context state "user satisfaction" and "technology acceptance". We consider five hypotheses in relation to the five alternatives (on the ordinal scale) for each QoE state. These are: "excellent (5)", "very good (4)", "good (3)", "fair (2)" and "poor (1)". We performed linear scale transformation such that the best alternative gets '1' and the worst alternative gets '0' using Eq. 4 i.e., $U(S_i^t = "excellent") = 1$ and $U(S_i^t = "poor") = 0$. Similarly, utility values: '0.75', '0.50' and '0.25', can be computed for other alternatives "very good", "good", "fair",

respectively. Eq. 4 can be written as:

$$U(a) = \frac{V(a) - V(a_{min})}{V(a_{max}) - V(a_{min})}$$
(4)

where U(a) represents the utility value of the alternatives. V(a) represents the value of the alternative. For example, "excellent" = 5, "very good" = 4, "good" = 3, "fair" = 2 and "poor" = 0. a_{max} and a_{min} represents the maximum (5) and minimum (1) values, respectively.

Once P(h) and U(a) are determined, the EU for each state is computed using Eq. 3. The EU will always fall within 0 and 1. This enables us to deal with uncertainty regarding QoE choices precisely as we show later in an example. The EU is mapped on the interval scale to determine the QoE of the states. For example, if $EU(S_{US}^t) = 0.8995$ (see Fig. 6), it means that $S_{US}^t = "excellent"$. The idea to map the utilities is based on the fact that users tend to give high rating with increasing number and lower rating with decreasing number as in the MEU principal we discussed earlier. These context states are then fused together to determine the overall situation of QoE (R_{QoE}^t) of users which is a global utility.

We consider a global utility which comprises several QoE states such as "user satisfaction" and "technology acceptance". There can be N QoE states affecting overall QoE of users. Thus, to determine the overall QoE, we consider a global utility GU(QoE) which is computed using the linear weighted sum of multiple QoE states. It can be written as:

$$GU(QoE) = \sum_{i=1}^{\kappa} w_i * EU\left(S_i^t\right)$$
(5)

 $EU(S_i^t)$ is computed using Eq. 3 and w_i are the weights associated with individual QoE states where $\sum_{i=1}^k w_i = 1$. Once the *GU* is computed, it is mapped onto the interval scale as shown in Fig. 6. to select the hypothesis, *a*. This hypothesis is the overall QoE situation (R_{QoE}^t). Algorithm 1 depicts the process of QoE measurement.

Once we have measured QoE based on the collected user ratings in a dataset (*trainingset*_{initial}), these QoE values are added as a new column in the dataset. We now call this dataset (*trainingset*_{append}) which can be used to create a new BN for QoE prediction on-the-fly as presented the next section.

3.3 Quality of Experience Learning and Prediction

For QoE prediction, we again considered BNs which involves learning the model parameters based on the collected dataset through subjective, simulation or experimental studies. In case of laboratory environments, users are usually guided through experiments carefully. Thus, any chance of missing information can be kept minimal. On the other hand, in case of user tests conducted in realistic user environments, users may or may not give ratings at all times. This makes it necessary for any methodology to deal with such cases. We deal with such missing and uncertain information by considering the use of expectation maximization (EM) algorithm [36] for our BNs parameter learning.

In EM algorithm [36], there are two steps: E-step or the expectation step and and M-step or the maximization step.

ALGORITHM 1: Algorithm to measure quality of experience using CaQoEM.

Input: *trainingset*_{initial}, context attributes (a_n^t) , QoE states (S_n^t) , and a bipolar interval scale. **Output**: QoE rating (R_{OOE}^t) .

- 2 Create/learn a Bayesian network (BN1) using domain knowledge or the initial training set (*trainingset_{initial}*).
- 3 Infer the beliefs of the BN and determine hypotheses *h* ∈ *H* for each QoE state (S_i^t) .
- ⁴ Calculate the expected utility of each QoE state $EU(S_i^t)$ using Eq. 3.
- 5 Map the $EU(S_i^t)$ onto the interval scale and determine QoE for that S_i^t .
- 6 Calculate the overall QoE based a global utility (*GU*(*QoE*)) based on Eq. 5.
- ⁷ Map the GU(QoE) on the interval scale and determine the overall QoE situation (R_{OoE}^t).
- s Record the R_{QoE}^t and append it to *trainingset*_{initial} training set (as a new column) and call is *trainingset*_{append}.
- 9 End

In case of complete data, it is equivalent to the maximum likelihood estimation (MLE). However, in case of missing data, it is a generalization of MLE. The basic idea is to find parameters (Θ) that maximizes the $logP(x; \Theta)$ of the data that is observed such as a_{delay}^t , a_{jitter}^t , $a_{location}^t$ and $a_{packetlosses}^t$. In case of the MLE, one can expect to find a global optima. However, in case of missing data, we can only find local optima. EM divides the problem into several sub-problems such that local optima of $logP(x; \Theta)$ can be found. In the E-step, for each sub-problem, there is a function f_t for which optimization is done. In case of M-step, a new parameter set (Θ^{t+1}) is considered that maximizes f_t . This is done iteratively until convergence. With our datasets, this was achieved in a few milliseconds (discussed in Section 4).

As shown in algorithms 1 and 2, once QoE is measured in terms of R_{QoE}^t , we append these values to the previous training set (*trainingset*_{initial}) and call it *trainingset*_{append}. Then a new BN, called BN_2 is created and is learnt via *trainingset*_{append}. BN_2 can now be used to predict new QoE ratings on new unseen data. In Section 4.3, we show that even simpler BNs can achieve an overall accuracy of 98.93% in multiple diverse network conditions which are typical in mobile computing environments.

4 **RESULTS ANALYSIS**

To validate CaQoEM, we considered a VoIP application and performed the following tasks: 1) we developed a number of case studies using the GeNIe 2.0 platform [1]; 2) we conducted user tests using a real-world prototype implementation; and 3) we performed extensive simulations studies using the OPNET simulator [3]. For results analysis, we considered the BayesNet Toolbox [28] for MATLAB and Weka data mining toolkit [15]. From the previous sections, we conclude that there can be complex and inter-dependent **ALGORITHM 2:** Quality of experience prediction using CaQoEM.

Input: *trainingset*_{append}. **Output**: QoE rating (R_{OOF}^t) .

- 1 Begin:
- 2 Create and learn a Bayesian network (*BN*₂) from the *trainingset_{append}* dataset.
- ³ Enter evidence to infer and predict the overall QoE (R_{QoE}^t) .
- 4 End

TABLE 2 Context Attributes (a_n^t) and Their Ranges

a_n^t	value range
codec	ITU-T G.711 and ITU-T G.729
packet losses	good (0%-5%), fair (5%-7%) and
	poor (7%-10%)
delay	good (0ms-150ms), fair (150
	ms-300ms) and poor (>300ms)
location	home, tram stop, office

relationships between several context attributes and QoE parameters. These relationships are very difficult to determine analytically or mathematically. In this section, we derive such relationships between different QoE parameters such as "user satisfaction" and "technology acceptance" and several context attributes as shown in Table 2. We argue that our model is flexible to encompass a wide range of QoE parameters and context attributes to correctly measure and predict users' QoE.

4.1 QoE Measurement for VoIP Applications in Mobile and Pervasive Computing Scenarios

4.1.1 Results Analysis Using Synthetic Data and Case Studies

In this section, we show how CaQoEM can be used for efficient QoE measurement in both laboratory and realistic user environments. We argue that CaQoEM is extensible as it incorporates several QoE parameters and context attributes to measure and predict users' QoE. We validate our approach by considering the MOS computed using the ITU-T E-Model [14] recommendation to determine state "user satisfaction" (S_{TA}^t) as another QoE parameter to determine the overall QoE situation (R_{QoE}^t). We define "technology acceptance" (TA) in relation to "user satisfaction"(US) by using the assumption that if a user is satisfied, he/she might be willing to accept the technology.

We use both US and TA to show how different QoE parameters can be used together to determine the overall QoE of users. Indeed, using CaQoEM any other QoE parameter can also be considered. The relationships between QoE parameters and context attributes are defined in a BN in the form of chance nodes and their probabilities are set using the conditional probability tables (CPTs). CPTs can be defined and set in several ways. They can be set using the domain knowledge provided by experts

¹ Begin:



Fig. 7. Performance of ITU-T G.711 and ITU-T G.729 codecs under different packet losses and delay.

and/or learnt through the data collected from user studies or experimentation [19], [36]. For example, data collected from user studies can be used to learn the CPTs. On the other hand, if experts want to add their expertise, they can manipulate the CPTs accordingly (by setting the probabilities) and can infer several hypotheses to reach a single (or multiple) conclusion(s) which is not possible in other QoE measurement techniques such as [10], [14], [24], [41]. This makes BNs to be a very flexible and powerful tool for QoE analysis, measurement and prediction.

Fig. 7 shows the motivation behind selecting particular values and ranges for codecs, packet loss and delay to determine S_{US}^t . This graph was derived from the ITU-T G.107 [14] and ITU-T G.113 [5] recommendations in relation to ITU-T G.711 (shown as solid lines) and ITU-T G.729 codecs (shown as dashed lines). Based on these observations, we defined the ranges for all the context attributes (see Table 2) of our BN. We discretized QoS parameters such as packet loss and delay into three different states such as "good", "fair" and "poor" and considered three different possibilities for location i.e., "home", "tram stop" and "office". From Fig. 7, we conclude that ITU-T G.711 codec performs better at 0% packet loss than the ITU-T G.729 codec but overall, ITU-T

G.729 codec performs significantly better even in case of "good" to "fair" packet loss and delay as compared to ITU-T G.711 codec by imposing lower bandwidth requirements. This is also true for the cases when "fair" to "poor" packet loss and delay are observed. For both codecs, bends around "fair" delay suggests that MOS further degrades rapidly from this point onwards due to increase in delay. Based on these observations, we set the CPTs.

As mentioned previously, experts can add domain knowledge in a BN to study several hypotheses. Thus, we further manipulated the CPTs to include the effects of location on the overall QoE based on the fact that a change in users' location affects their QoE [21]. In reality, users' location information can be gathered using GPS sensors on their mobile phone alongside their subjective ratings to realistically determine their QoE [12]. We intend to study affects of location via experimentation as a future work. For the scope of this paper, we assume that users are more comfortable at home rather than in their office or tram stops due to stress or background noise which might affect their QoE. This was achieved by assigning higher probability to home followed by tram stop and the office to maximize the overall QoE (R_{OoE}^t).

In order to measure users' QoE using CaQoEM, context states S_{SA}^t and S_{TA}^t are first inferred probabilistically and then fused together to determine a single scalar value or global utility GU(QoE) to determine the overall QoE situation (R_{QoE}^t) as shown in the algorithm 1. We considered Clustering algorithm [36] to infer context states taking values between '1' and '5';'1' means "poor", '2' means "fair", '3' means "good", '4' means "very good" and '5' means "excellent". These values can help the stakeholders in understanding hypotheses such as "user satisfaction is fair" or "user is not willing to accept the technology". In this case, S_{US}^t can be inferred by setting the evidence variables (e) such as $S_{US}^t = P(user satisfaction|location =$ "home", <math>delay = "good", packet loss = "good").

We conducted several experiments as shown in Table 3. For the scope of this paper, we present eight cases from a

TABLE 3

Observed Context Parameters with Probabilities. Final QoE Utility Value GU(QoE) is Calculated and is Mapped onto the Bipolar Interval Scale to Determine the Overall QoE Situation (R_{OoE}^t)

Exp	Codecs (probability)	Packet loss	delay	location (probability)	GU(QoE)	R^t_{QoE}
No.		(probability)	(probability)			
1.	ITU-T G.711 (1)	good (1)	good (1)	tram stop (1)	0.7375	very good
2.	ITU-T G.711 (1)	poor (1)	poor (1)	tram stop (1)	0	poor
3.	ITU-T G.729 (1)	good (0.90), fair	good (0.60), fair	home (1)	0.6768	very good
		(0.10), poor (0)	(0.25),			
			poor (0.15)			
4.	ITU-T G.729 (1)	fair (1)	fair (1)	home (1)	0.5000	good
5.	ITU-T G.729 (1)	poor (1)	poor (1)	tram stop (1)	0.0062	poor
6.	ITU-T G.711 (1)	good (1)	good (1)	home (1)	0.9750	excellent
7.	ITU-T G.711 (1)	good (0.20), fair	good (0.30), fair	tram stop(1)	0.1847	fair
		(0.60),	(0.30),	_		
		poor (0.20)	poor (0.40)			
8.	ITU-T G.729 (1)	good (0.80), fair	good (0.40), fair	office (1)	0.5616	good
		(0.20), poor (0)	(0.30),			
			poor (0.30)			



Fig. 8. Screenshot of CaQoEM implemented in GeNIe platform.

total of fourteen which show the capabilities of our proposed approach. Consider experiment 1, where a user is standing at the tram stop during peak morning hours. He/she receives a call from one of their colleagues. The mobile node gathers statistics regarding QoS context parameters. For instance, "good" values for packet losses and "good" values for delay were observed. To determine the overall QoE situation (R_{OoE}^t) , S_{US}^t and S_{TA}^t were inferred. Our algorithm predicted "very good" and "excellent" for S_{US}^t with probabilities 0.10 and 0.90. In case of S_{TA}^t , the algorithm predicted "very good", "good" and "fair" with probabilities 0.25, 0.50 and 0.25, respectively. We considered Eqs. 2 to 5 to determine R_{OoE}^{t} . For context state, "user satisfaction", the expected utility $EU(S_{US}^t)$ was computed as 0.9750. As $EU(S_{US}^t)$ was 0.9750 and in the range [0.8751:1], the S_{US}^t was "excellent" (see Fig. 6). To validate that it is a correct result, there are two conditions: 1) the value should lie between 0.75 ("very good") and 1 ("excellent") as the beliefs for only two outcomes were inferred; and 2) it should be very close to 1 as 90% of users selected "excellent" and 10% selected "very good" as an outcome i.e., most (90%) of the users are inclined towards "excellent" or '1' (on the interval scale. See Fig. 6). As $EU(S_{US}^t)$ is in the correct range and is very close to '1', S_{US}^t as "excellent" is a correct result.

Similarly, for context state, technology acceptance, $EU(S_{TA}^t)$ was computed as 0.5000. Once we determined all QoE states, the overall QoE situation (R_{QoE}^t) was determined. We used Eq. 5 for this. The idea is to have a single scalar value that can be used by the stakeholders to make informed decisions regarding problems such as customer churn management, network optimization and codec engineering. Eq. 5 considers a linear weighted sum of each QoE parameter to compute the global utility that can be mapped easily on the interval scale. For the sake of brevity, we consider equal weights of 0.50 for each QoE state. Thus, the global utility is calculated as 0.7375. After GU(QoE) is calculated, it is mapped onto the interval scale to determine R_{QoE}^t . In this particular example, it is determined as "very good". Fig. 8 shows

the screenshot of this example implemented in the GeNIe platform [1].

Consider experiment 2 (shown in Table 3), where a user uses a VoIP application at a "tram stop" and the mobile node observes "poor" packet loss and delay values. In this case, the overall QoE was "poor" with GU(QoE) calculated as '0' which was correctly determined by the model as all the users gave "poor" ratings for both S_{IIS}^t and S_{TA}^{t} . Hence the EU in both cases was 0 leading to GU(QoE) as 0. This is an extreme case which shows the worst QoE is correctly mapped to 0. Similarly, consider experiment 3 where a user is sitting at his/her home and is relaxed. The observed packet loss and delay were mostly "good" while using the ITU-T G.729 codec. In this case, the R_{OoF}^{t} was measured as "very good". Similarly, we conducted several other experiments as shown in Table 3 to evaluate our proposed approach. These experiments validate that CaQoEM correctly measures users' QoE. The correctness of the results was demonstrated by first considering the extreme cases such as case 6 and case 2 where the maximum and minimum GU(QoE) were derived in case of best and worst QoS conditions at different user locations. We then considered other cases to infer and measure QoE with different parameter settings and outcomes.

4.1.2 Results Analysis Based on User Studies/Subjective Tests Method

We recruited 29 participants from diverse backgrounds to assess VoIP call quality for two different codecs (ITU-T G.722 wideband codec (WB) and ITU-T G.711 narrowband (NB) codec) under several network impairments which are typical in mobile computing scenarios. These network impairments include: 0%, 5%, 10% and 20% packet losses and horizontal handoffs. All the participants were fluent in English language. There were 26 male and 3 female participants working at the Department of Computer Science, Electrical and Space Engineering at Luleå University of Technology, Sweden. The choice of the number of participants for the subjective study was based on studies presented in the state-of-the-art [9], [33], [41]. For example, Wu et al. [41] considered sixteen users for subjective and objective tests. Pinson et al. [33] suggests that subjects (users) greater than 24 are sufficient for conducting subjective tests based on ordinal scale.

The average age of all participants was 31.24 years. The age of users may or may not affect their QoE depending on whether 'age' is included as a context attribute in a BN model. To verify this hypothesis, subjective tests involving a number of users with different age would be required. The BN model would then be learnt and conclusions regarding QoE may be drawn. For instance, if the parent node in a BN, say 'age' affects the probability distribution of the child node 'QoE'. It means age affects users' QoE, otherwise it does not. In the user studies, we do not consider 'age' as a context attribute. However, CaQoEM can easily incorporate 'age' as a context attribute and others mentioned in Table 1 to model, measure and predict users' QoE.



Fig. 9. Testbed setup for user experiments. Netem enabled router is used to introduce network impairments such as packet losses.

We created a simple web interface where users could login, listen to the speech samples and then give their ratings on the ordinal scale of 1 to 5. On this scale, '1' means "poor", '2' means "fair", '3' means "good", '4' means "very good" and '5' means "excellent". For this study, we considered "office context" where the participants were asked to perform the test in their office using the web interface.

Testbed Setup

To assess VoIP call quality, we developed a VoIP-based testbed comprising of two PCs and a Netem [2] enabled router as shown in Fig. 9. We used speech samples available from the ITU-T P.50 database [6] and used PJSIP VoIP application [4] on the correspondent node (CN) and the mobile node (MN). PJSIP enables automatic voice play and recording functionality and supports numerous NB and WB codec. To prepare speech samples for user tests, we introduced impairments such as 0%, 5%, 10% and 20% packet losses and horizontal handoffs using Netem daemon running on the router during an active VoIP session. Netem can emulate delay, packet loss, packet duplication and packet reordering. In our testbed, we introduced random packet losses directly using Netem's commands and by not emulating the packet losses through bandwidth restriction between MN and the CN. This is because Netem is capable of generating realistic packet losses in a controlled manner thereby eliminating the need for packet loss generation without bandwidth restriction.

As can be seen in Fig. 9, VoIP calls were initiated from a CN towards the MN. During the call, packet losses (PL) were randomly generated. Further, horizontal handoffs were introduced in the middle of a VoIP call by varying the Netem setup. The calls were then recorded at the MN in the form of degraded speech samples. Once all speech samples were generated for both codecs related to all network impairments, we then asked the participants to listen to these voice samples via a simple web interface. The ratings provided by the participants were stored in the MySql database.

Results Analysis

Fig. 10 presents the results related to user studies for VoIP application employing a NB and a WB codec. In this figure, cases 1 to 5 are for the NB codec and cases 6 to 10 represent the results for WB codec. Cases 1 and 6 with 0% PL provide best user satisfaction followed by cases related to 5% PL, 10% PL, horizontal handoff and 20% PL, respectively. In all cases, WB codec gives best user satisfaction. In particular, for the best case scenario with 0% PL, most of the users selected the maximum achievable score of '5' with probability 0.59. However, in case of the NB codec with 0%



Fig. 10. Distribution of user preferences for narrowband and wideband codecs with several network impairments. (a) ITU-T G.711 narrowband (NB) codec. (b) ITU-T G.722 wideband (WB) codec.

packet loss, only 3% of the users selected '5' while 41% of the users selected '4'.

After all user ratings were collected, we then trained a BN to measure QoE using CaQoEM. An important aspect to consider for QoE measurement using BNs is whether to select an outcome based on maximum a posteriori (MAP). For example, consider a case where we have 10 outcomes in which six users gave '5' to call quality and four users gave '4' to call quality. From probabilistic point of view, the MAP based outcome is '5' with probability 0.60. This assumption will be incorrect especially when the data is scarce. This will be true for methods proposed in [24], [35] where the authors considered MAP based QoE measurement and prediction. To deal with this problem, we proposed a decision-theoretic approach where rather than selecting an outcome with MAP, we consider a QoE value that considers all the ratings taken together i.e., we find an alternative that "best describes" the underlying QoE ratings. This is done via considering all the QoE ratings together to determine a single scalar value rather than just selecting the most likely outcome based on highest probability. This scalar value is then mapped to the interval scale to determine a final outcome. This outcome is then

TABLE 4 Overall QoE Determined by CaQoEM for User Studies

Cases	GU(QoE)	R^t_{QoE}
1.) ITU-T G711 with 0% PL	0.5690	good
2.) ITU-T G711 with 5% PL	0.3707	fair
3.) ITU-T G711 with 10% PL	0.2155	fair
4.) ITU-T G711 with 20% PL	0.0948	poor
5.) ITU-T G711 with handoff	0.3362	fair
6.) ITU-T G722 with 0% PL	0.8707	very good
7.) ITU-T G722 with 5% PL	0.5431	good
8.) ITU-T G722 with 10% PL	0.3276	fair
9.) ITU-T G722 with 20% PL	0.1207	poor
10.) ITU-T G722 with handoff	0.5172	good

associated with the collected dataset to perform prediction as shown in the next section.

Fig. 10 shows the probability distribution for all cases. As shown in Fig. 10(a), for case 1, the MAP based QoE outcome would be "very good" or '4'. For case 2, it would be "fair" ('2'), followed by "fair" ('2'), "poor" ('1') and "good" ('3') for cases 3, 4 and 5, respectively. Compared to random neural networks and decision trees based methods [24], [35], we assert that selecting MAP based QoE may not vield correct results as there are other outcomes which are also very close to each other. For example, in case 2 (Fig. 10(a)), alternatives "very good" and "good" are very close with probabilities 0.41 and 0.45. Also, alternatives "excellent" and "poor" are equally likely. Thus, to deal with uncertainty and imprecision in QoE measurement, it therefore makes sense to consider all user ratings together to determine a single most likely alternative by considering uncertainty regarding underlying probability distribution.

In case of ITU-T Recommendations such as ITU-T G.107 [14] and P.800 [30], mean of user ratings is taken to determine QoE. As discussed in the previous section, these methods may not be appropriate as the users have no idea between the relative distance between all the outcomes (considering ordinal scale) as their ratings are subjective rather than objective in nature [16]. To tackle the problem of mapping subjective outcomes to objective outcomes, we consider a bipolar scale and linear scale transformation method where the distance between the outcomes is determined by the experts. In most cases, the distance between the outcomes are considered equal. Thus, in this paper we use this method to measure users' QoE and show that it reaches towards selecting correct alternatives. In user studies, we considered packet losses to determine users' QoE in the form of user satisfaction. Thus, we aim to find $EU(S_{OoF}^t)$ as in Eq. 3.

Consider case 1 (Fig. 10(a)) which is the best case scenario for NB codec with 0% PL. In this case, most of the users selected '4' as the outcome with probability 0.41 followed by selecting the outcomes '3', '2' and '1' with probabilities 0.34, 0.21 and 0. Surprisingly, only 4% of the users selected '5' as the outcome. It is interesting to note that the user ratings in this case were widely distributed between '5' and '2'. Thus, the GU(QoE) was calculated as 0.5690. Mapping this value on the interval scale (shown in Fig. 6), the overall QoE (R_{QoE}^t) was determined as "good". In case 8 (Fig. 10(b)), we observed that 59% of the users selected '2' as the outcome. 31% users selected '3' followed by 7% selecting '1' and 3% selecting '4' as the outcome. The GU(QoE) calculated in this case was 0.3276 which means overall QoE is "fair".

Consider case 9 (Fig. 10(b)) as the worst case scenario for the WB codec. In this case, most of the users selected '1' as the outcome with probability 0.62 followed by '2' with probability 0.28 and '3' with probability 0.10. Thus, the GU(QoE) was calculated as 0.1207. As this value is in the range: [0:0.1250], the overall QoE was "poor". In this case, we can observe that 38% of the users gave ratings better than "poor". Thus, this value lies closer to the middle of "poor" and "fair" but is more weighted towards "poor". Thus, "poor" was correctly selected as the outcome. Similarly, we calculated QoE for all the cases for both NB and WB codecs as shown in Table 4.

It is worth mentioning here that the OoE in case of horizontal handoffs (cases 5 and 10) is relatively high for both NB and WB codecs and is equivalent to 5% PL for both codecs. This is mainly due to the fact that for our subjective tests we considered one-way tests similar to [26] where users only listened to the voice conversation. In case of twoway or conversational tests, we expect lower QoE ratings in case of horizontal tests. Based on the results presented in Table 4, we assert that packet losses upto 5% should be the limit for NB codec while for WB codec, this limit is about 10%. The results in this section validate that CaQoEM correctly measures users' OoE. The correctness of the results was again demonstrated by considering extreme cases such as cases 1, 6, 4 and 9 where maximum and minimum GU(QoE) values were correctly computed corresponding to the type of network impairment.

4.1.3 Discussion and Lessons Learnt

In this section, we demonstrated the QoE measurement capability of CaQoEM with case studies and subjective tests. We discussed that if the data is scarce or is plagued with variability and uncertainty, imprecision in QoE measurement results can occur. This can be due to wrong predictions made by any machine learning methods. CaQoEM solves this problem by finding the overall QoE situation (R_{OoE}^t) that best describes the underlying data by considering all the outcomes provided by users and selecting a single scalar value that determines the overall QoE. This is achieved by first inferring the QoE states probabilistically based on subjective user ratings. These subjective ratings are then quantified as expected utilities based on utility theory and the interval scale. Then each QoE state is mapped on the interval scale to select the most appropriate QoE rating (R_{OoE}^t) . We must stress here that using our method, stakeholders can use any type of scale. For example, a scale with range: [-5:+5] or a scale with range: [0:100] to determine users' QoE as a single scalar value by setting utilities according to their requirements.

In all, CaQoEM offers several advantages: 1) It provides a naturalistic way to define relationships between context attributes and QoE states for efficient context-aware QoE modelling and measurement. 2) It facilitates graceful context addition and removal to model, measure and predict users' QoE. 3) It easily incorporates domain/experts knowledge for QoE inference. 4) It handles uncertainty and imprecision regarding users' QoE ratings in an efficient and unbiased manner. 5) It maps several QoE states and context attributes to measure users' QoE on a single scale.

4.2 Evaluating QoE for VoIP Applications in Mobile and Pervasive Computing Scenarios

Previous section presented results related to QoE measurement. In this section, we study how QoE measured by CaQoEM can be used to perform prediction in mobile computing scenarios where mobility poses serious challenges to users' QoE due to impairments such as handoffs and wireless signal fading. In this section, we first study the impact of these impairments on VoIP application and then study how BNs can be used to predict users' QoE with high accuracy. We also discuss in detail two different types of



Fig. 11. Multiple scenarios which impact QoE related to multimedia applications such as VoIP. (a) Fading. (b) Vertical handoff. (c) Network congestion. (d) Hotspot traffic.

BNs that can be employed by the stakeholders to measure and predict users' QoE. We consider several scenarios such as wireless signal fading, varying traffic load, and vertical handoffs in IEEE 802.11 WLAN network environments as shown in Fig. 11.

We considered VoIP application that can use both ITU-T G.711 and ITU-T G.729 voice codecs. To determine "user satisfaction" (S_{US}^t) on the scale of 1 to 5, we considered the ITU-T E-Model [14]. Where '1' means "poor", '2' means "fair", '3' means "good", '4' means "very good" and '5' means "excellent", respectively. Indeed, our model competes directly against the ITU-T E-Model but in order to test the measurement and prediction capability of our proposed method under diverse network conditions, we find the output of ITU-T E-Model to be sufficient for our purposes as it consider non-linearity in delay and packet loss impairments to output the MOS. Further, using OPNET simulator [3], we can generate different datasets for several scenarios based on realistic simulation setup and assumptions. Using our studies, we gather that CaQoEM is both space and time efficient and can be used in scenarios where correct

QoE predictions are required on-the-fly. Using CaQoEM, the applications do not have to perform expensive calculations to obtain MOS using the ITU-T E-Model yet resulting in significant QoE prediction accuracy.

4.2.1 Wireless Signal Fading in IEEE 802.11g WLAN

In case of wireless signal fading in IEEE 802.11g WLAN, a mobile node (MN) moved out of the coverage area of an access point (AP) while having a voice conversation with the correspondent node (CN). In this scenario, we only wanted to determine the effects of mobility on voice conversation. Thus, we do not consider any background network traffic. The call length was assumed to be 30 seconds after which the MN moved out of the coverage area of the AP and the call was dropped. Fig. 11(a) shows the targeted scenario. As soon as the MN reached a certain signal to noise ratio (SNR) threshold (14dB), the MOS suddenly dropped from 3.6 to 1. The cause of the sudden drop in the MOS can be attributed to the sudden rise in network delay due to extremely low SNR values. The end-to-end delay increased

TABLE 5 Cumulative Results Related to All the Scenarios for ITU-T G.711 Voice Codec

Scenario	Avg. delay (ms)	Avg. PLR (%)	Avg jitter (ms)	Avg. E-Model MOS
Network congestion	414	0.03	0	2.06
Signal fading	93	0	0.06	3.59
Vertical handoff	108	10.05	0	2.75
Normal hotspot traffic	71	0	0	3.62

Scenario	Avg. delay (ms)	Avg. PLR (%)	Avg. jitter (ms)	Avg E-Model MOS
Network congestion	170	0	0	2.41
Signal fading	92	0	0.02	3.00
Vertical handoff	108	7.61	0	2.50
Normal hotspot traffic	70	0	0	3.01

 TABLE 6

 Cumulative Results Related to All the Scenarios for ITU-T G.729 Voice Codec

suddenly from 0.060 milliseconds (ms) approx. to 1 second (sec) approx for both codecs.

4.2.2 Vertical Handoffs in Vehicular Networks Using Mobile IPv6

We considered vertical handoffs between two WLAN networks at low vehicular speeds (10 km/hr). We used Mobile IPv6 [20] mobility management protocol. A vehicle carrying a MN moved from one WLAN network to another. The MN then went through handoffs (shown in Fig. 11(b)). In this case, the delay caused by handoff was quite low at 108 ms with average packet loss ratio of 10.05% and 7.61% for ITU-T G.711 and ITU-T G.729 codecs, respectively. These statistics are fairly typical in case of hard handoffs where connection from one network is lost so that a MN can attach itself to a new network. We assumed that handoffs occurs at the middle of a voice call. In reality, location of handoff within a call can have different impact in users' QoE [26].

4.2.3 IEEE 802.11b WLAN Network Congestion

To determine the effects of wireless network congestion on the voice quality, we used the IEEE 802.11b WLAN. We studied how MOS varies with an increase in network traffic load on the access point (AP). To saturate an AP, we set up three wireless nodes to generate background UDP traffic as shown in Fig. 11(c). The maximum achievable bit rate was approximately 5Mb/s after which the AP dropped all packets due to buffer overflow. All three nodes generated 175 packets per second towards the CN. Once the network reached the steady state, the MN initiated a call to the CN and calculated the required statistics. In this case, we set the call duration to 16 minutes and calculated the average statistics. From these statistics, we gather that in case of heavy UDP traffic, end-to-end delay causes MOS to vary. We observed that overall packet loss rate and jitter were close to 0 for both codecs. The average MOS for ITU-T G.711 codec was 2.06 and 2.41 for ITU-T G.711 codec, respectively.

4.2.4 Normal Hotspot Network Traffic

In this case we considered the IEEE 802.11g WLAN where the MN enjoyed near perfect voice call with extremely low delay and jitter. Fig. 11(d) shows the targeted scenario. Two MNs generated approximately 4 Mb/s of background UDP network traffic towards the CN. After sometime, a voice call was established between the CN and the MN. As the amount of traffic generated was not enough to cause congestion, the VoIP call enjoyed very low delays and close to zero jitter and packet losses. Hence, the average MOS calculated was greater than 3 for both codecs. Tables 5 and 6 show the results related to all the scenarios for both codecs.

4.3 Quality of Experience Prediction in Mobile Computing Scenarios

A BN for QoE measurement and prediction can be created in many ways. We can create a BN by creating 'arcs' between all context attributes and QoE parameters. We can then learn the BN parameters from the collected data. Using Eq. 2, any query in the BN can be answered in a straightforward manner. By using inference algorithms such as variable elimination and junction tree [36], we can determine hypotheses about users' QoE efficiently. Fig. 12(a) and (b) shows two BNs for QoE prediction based on noisy-OR model and naive Bayes' model [19], [36]. One of the simplest BN is the naive Bayes' network model (NBN) as shown in Fig. 12(b) where all the observations or context attributes are conditionally independent given an outcome such as, QoE is "excellent". Thus, we have a generative model in the form of:

$$P(R_{QoE}^t|a_n^t) = P(R_{QoE}^t) \prod_n P(a_n^t|R_{QoE}^t)$$
(6)

In some cases, the independence assumption regarding context attributes to determine QoE may be too strict. In this model, the observations are assumed to the discrete i.e., a_{delay}^t , a_{jitter}^t and $a_{packetlosses}^t$ are discretized into finite states. Another form of NBN is the Gaussian naive Bayes (GNB) which assumes the observations or context attributes are represented by Gaussian distribution. This model can handle continuous observations directly without the need for discretization.

Another simple model as shown in Fig. 12(a) is the noisy-OR model or the discriminative model. This model is used to predict the QoE state directly from the given observations i.e., $P(R_{QoE}^t|a_n^t)$. For a hybrid BNs (HBN) with continuous and discrete nodes, one has to use the logistic or softmax distribution to perform inference. The logistic function is written as:

$$P(R = r_i | a_1^t, \dots, a_n^t) = \frac{exp(b^i + \sum_{l=1}^n w_l^i a_l^t)}{\sum_{j=1}^m exp(b^j + \sum_{l=1}^n w_l^j a_l^t)}$$
(7)



Fig. 12. Bayesian networks for QoE measurement and prediction. (a) Noisy-OR model. (b) Naive Bayes model.

delay

jitter



nearby

heart rate

screen size

mood

location

Fig. 13. Possible complex relationships between QoE parameters and context attributes and among QoE parameters themselves. Grey ovals represent QoE parameters and white ovals represent context attributes. Attributes such as "bandwidth," "mood," and "people nearby" are secondary context attributes which are determined based on primary context attributes such as "blood pressure" and "location."

blood

pressure

packet loss

bandwidth

Eq. 7 defines a softmax model, which is similar to performing multinomial logistic regression when continuous nodes (a_n^t) are the parents of the discrete nodes (R). An important property here is that in a softmax model, stakeholders can add domain knowledge which is not possible in simple logistic regression. In this model, parent nodes are defined usually as Gaussian distribution for continuous attributes such as $a_{bandwidth}^t$ and conditional probability tables (CPT) are defined for discrete nodes such as $a_{location}^t$. The CPDs of the children nodes such as S_{US}^t is defined as the softmax nodes as S_{US}^t is measured on the ordinal scale. There are various other ways to construct a BN which is out of the scope of this paper. Indeed, we will show that how these simple models can predict QoE very accurately.

For each aforementioned scenario, we collected results based on 125 simulation runs with different random seeds in OPNET network simulator [3]. We obtained 2 datasets related to both ITU-T G.711 (dataset 1) and ITU-T G.729 (datset 2) codecs. In both datasets, we had 500 training cases each representing results related to diverse network conditions. Tables 7 and 8 shows the statistics related to all the parameters in both datasets. We then trained our BNs to predict users' QoE. For training the model parameters, we use the expectation maximization (EM) algorithm [36]. We used Bayes Net Toolbox for MATLAB (BNT) [28] for analyzing the HBN and used Weka data mining software [15] for analyzing NBN and GNB models. Both datasets were complete with no missing values. Hence, EM reduced to maximum likelihood estimation (MLE) as in the

TABLE 7 Statistics Related to All the Parameters in the Dataset 1

a_n^t	Min. value	Max. value	Mean	Std. dev
a _{delay}	68 ms	671 ms	172 ms	146 ms
a_{jitter}	0 ms	3 ms	1 ms	1 ms
$a_{packetloss}$	2.5 %	25.40%	2.5%	4.6%
S_{US}	1	4	-	-

TABLE 8 Statistics Related to All the Parameters in the Dataset 2

a_n^t	Min.	Max.	Mean	Std.
	value	value		dev
a_{delay}	67	374	109 ms	41
	ms	ms		ms
a_{jitter}	-1 ms	3 ms	1 ms	1 ms
$a_{packetloss}$	0 %	20.5%	1.9%	3.7%
S_{US}	1	3	-	-

supervised learning case. As we are dealing with learning and prediction in case of BNs, the best approach for their evaluation is to perform cross validation where the idea is to estimate models prediction accuracy by keeping aside some fraction of the training data for learning. The remaining data is used as the test data for which hypothesis of the states are inferred [36]. We used 10-fold cross validation for checking models prediction accuracy.

Tables 9 and 10 show the prediction accuracy for both datasets. As can be observed, all three types of BNs give excellent prediction accuracy based on 10-fold cross validation. Even these simple BNs (Fig. 12(a) and (b)) were able to correctly predict users' QoE with an overall average accuracy of 98.93%. In dataset 1, both generative (NBN and GNB) and discriminative (HBN) BNs perform well with all three BNs achieving a prediction accuracy greater than 98%. In case of dataset 2, HBN performs marginally better than GNB and NBN with prediction accuracy of 99.80%. NBN and GNB both achieved prediction accuracy of 98.60%, respectively. A prediction accuracy of greater than 95% was achieved in all cases which is an excellent result.

4.3.1 Discussion and Lessons Learnt

From results presented in this section, we conclude that BNs can be used for accurate QoE prediction. From implementation perspective, both discriminative and generative BNs work well on our dataset. However, generative or discriminative BNs may outperform each other in different cases. In such a scenario, a BN should be chosen which provides better prediction accuracy. Applications can be developed with ease, that take the available datasets as input and automatically output the correct QoE measurements by selecting an optimal BN, for example, using Weka [15] or GeNIe APIs [1]. It is worth noting that more objective and subjective parameters can be added and deleted as the per the stakeholders' requirements and BNs can easily accommodate such changes while maintaining low complexity and higher accuracy. These BNs can

TABLE 9 Prediction Accuracy of NBN, GNB, and HBN for Voice Quality Measurement for Dataset 1

Bayesian network type	prediction accuracy (%)
Naive Bayes Network (NBN)	99
Gaussian Naive Bayes Network (GNB)	98.40
Hybrid Bayesian Network (HBN)	99.20

TABLE 10 Prediction Accuracy of NBN, GNB, and HBN for Voice Quality Measurement for Dataset 2

Bayesian network type	prediction accuracy (%)
Naive Bayes Network (NBN)	98.60
Gaussian Naive Bayes Network (GNB)	98.60
Hybrid Bayesian Network (HBN)	99.80

also be applied directly to real-time applications and protocols for dynamic network selection and codec change. Inference time is less than 50 milliseconds for NBN and GNB and around 120 milliseconds for HBN which also include learning the model parameters.

5 CONCLUSION AND FUTURE WORK

This paper presented a novel context-aware approach called CaQoEM for context-aware quality of experience (QoE) modelling, measurement and prediction. CaQoEM incorporates several context attributes and QoE parameters to measure and predict users' QoE under uncertainty using Bayesian networks, utility theory and a bipolar scale. Our results comprising of case studies and user evaluations show that CaQoEM can efficiently measure QoE based on multiple QoE parameters and context attributes. Using simulation studies, we study VoIP applications under multiple diverse network conditions such as vertical handoffs, network congestion and wireless signal fading which are prevalent in mobile computing systems. We show that CaQoEM efficiently learns and predicts QoE using BNs and achieves an overall prediction accuracy of 98.93%. We believe CaQoEM can benefit stakeholders such as network operators, codec engineers or application designers who are interested in modelling, measuring and predicting users' QoE. In the future, we will incorporate CaQoEM for network churn management. Further, we will conduct several longitudinal users studies using CaQoEM.

REFERENCES

- [1] *Genie Software Package* [Online]. Available: http://genie.sis.pitt.edu/about.html, accessed Nov. 2, 2011.
- [2] Netem-Network Emulation For Linux [Online]. Available: http://netgroup.uniroma2.it/twiki/bin/view.cgi/main/netemclg, accessed Nov. 2, 2011.
- [3] Opnet Network Simulator. [Online]. Available: http://www.opnet.com, accessed May 5, 2011.
- [4] Pjsip-Open Source Sip Stack And Media Stack For Presence, Im/Instant Messaging, And Multimedia Communication. [Online]. Available: www.pjsip.org, accessed Nov. 2, 2011.
- [5] Provisional Planning Values for the Equipment Impairment Factor IE, ITU-T Recommendation G.113A–Appendix I, 1998.
- [6] Artificial Voices, ITU-T Recommendation P.50, 1998.
- [7] C. Bettini et al., "A survey of context modelling and reasoning techniques," Pervasive Mobile Comput., vol. 6, no. 2, pp. 161–180, 2010.
- [8] P. Brooks and B. Hestnes, "User measures of quality of experience: Why being objective and quantitative is important," *Netw. IEEE*, vol. 24, no. 2, pp. 8–13, Mar.–Apr. 2010.
- [9] K. Chen, C. Tu, and W. Xiao, "OneClick: A framework for measuring network quality of experience," in *Proc. INFOCOM IEEE*, Rio de Janeiro, Brazil, 2009, pp. 702–710.
- [10] K. T. Chen, C. Y. Huang, P. Huang, and C. L. Lei, "Quantifying skype user satisfaction," in *Proc. SIGCOMM*, Pisa, Italy, 2006.

- [11] L. De Marez and K. De Moor, "The challenge of user-and QoE-centric research and product development in today's ICTenvironment," *Observatorio* (*OBS**), vol. 1, no. 3, pp. 1–22, 2007.
- [12] K. De Moor et al., "Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting," Mob. Netw. Appl., vol. 15, pp. 378–391, Jun. 2010.
- [13] A. K. Dey and G. D. Abowd, "Toward a better understanding of context and context-awareness," College Comput., Georgia Institute Tech., Atlanta, GA, USA, Tech. Rep. git-gvu-99-22 [Online]. Available: (try) / (try ac astersh edu (ryth (ryth / 1000 / 00 22 mdf))

ftp://ftp.cc.gatech.edu/pub/gvu/tr/1999/99-22.pdf

- [14] The E-model: A Computational Model for Use in Transmission Planning, ITU-T Recommendation G.107, 2008.
- [15] M. Hall et al., "The weka data mining software: An update," SIGKDD Explor. Newslett., vol. 11, pp. 10–18, Nov. 2009.
- [16] C. L. Hwang and K. P. Yoon, Multiple Attribute Decision-Making: Methods and Applications. Berlin, Germany: Springer-Verlag, 1981.
- [17] R. Jain, "Quality of experience," *Multimedia IEEE*, vol. 11, no. 1, pp. 96–95, Jan.–Mar. 2004.
- [18] L. Janowski and Z. Papir, "Modeling subjective tests of quality of experience with a generalized linear model," in *Proc. QoMEx*, San Diego, CA, USA, 2009, pp. 35–40.
- [19] F. V. Jensen, Bayesian Networks and Decision Graphs. New York, NY, USA: Springer, 2001.
- [20] D. B. Johnson, C. E. Perkins, and J. Arkko. (Jun. 2004). Mobility Support in ipv6 (RFC 3775) [Online]. Available: http://www.ietf.org/rfc/rfc3775.txt
- [21] S. Jumisko-Pyykkö and M. M. Hannuksela, "Does context matter in quality evaluation of mobile television?" in *Proc. 10th MobileHCI 2008*, Amsterdam, The Netherlands, pp. 63–72.
- [22] K. Kilkki, "Quality of experience in communications ecosystem," J. UCS, vol. 14, no. 5, pp. 615–624, Mar. 2008.
- [23] K. U. R. Laghari and K. Connelly, "Toward total quality of experience: A QoE model in a communication ecosystem," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 58–65, Apr. 2012.
- [24] V. Menkovski, A. Oredope, A. Liotta, and A. C. Sánchez, "Predicting quality of experience in multimedia streaming," in *Proc. 7th MoMM*, New York, NY, USA, 2009. pp. 52–59.
- [25] K. Mitra, C. Åhlund, and A. Zaslavsky, "A decision-theoretic approach for quality-of-experience measurement and prediction," in *Proc. ICME IEEE*, Barcelona, Spain, Jul. 2011, pp. 1–4.
- [26] S. Moller, M. Waltermann, B. Lewcio, N. Kirschnick, and P. Vidales, "Speech quality while roaming in next generation networks," in *Proc. ICC IEEE*, Dresden, Germany, 2009, pp. 1–5.
- [27] M. Mu, A. Mauthe, G. Tyson, and E. Cerqueira, "Statistical analysis of ordinal user opinion scores," in *Proc. CCNC IEEE*, Las Vegas, NV, USA, Jan. 2012, pp. 331–336.
- [28] K. Murphy, "The Bayes net toolbox for matlab," in Proc. Comput. Sci. Statist., 2001.
- [29] (2011, Sep. 10) Nokia. Quality Of Experience (QoE) Of Mobile Services: Can It Be Measured And Improved? [Online]. Available: http://www.nokia.com/nokia_com_1/operators/downloads/ nokia_services/whitepaper_qoe_net.pdf
- [30] Methods for Subjective Determination of Transmission Quality, ITU-T Recommendation P.800, 1996.
- [31] A. Padovitz, S. W. Loke, A. Zaslavsky, B. Burg, and C. Bartolini, "An approach to data fusion for context awareness," in *Proc. 5th CONTEXT*, Paris, France, 2005.
- [32] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.
- [33] M. H. Pinson *et al.*, "The influence of subjects and environment on audiovisual subjective tests: An international study," *IEEE J. Sel. Topics Sig. Process.*, vol. 6, no. 6, pp. 640–651, Oct. 2012.
- [34] K. K. Rachuri *et al.*, "EmotionSense: A mobile phones based adaptive platform for experimental social psychology research," in *Proc. 12th ACM Ubicomp*, Copenhagen, Denmark, 2010, pp. 281–290.
- [35] G. Rubino, P. Tirilly, and M. Varela, "Evaluating users satisfaction in packet networks using random neural networks," in *Proc. ICANN*, Berlin, Germany, 2006, pp. 303–312, LNCS 4131.
- [36] S. Russel and P. Norvig, Artificial Intelligence: A Modern Apporach, 2nd ed. Beijing, China: Tsinghua University Press, 2006.
- [37] J. Sun, "Football on mobile phones: Algorithms, architectures and quality of experience in streaming video," Ph.D. thesis, Umeå Univ., Umeå, Sweden, 2006.

IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 14, NO. 5, MAY 2015

- [38] A. Takahashi, H. Yoshino, and N. Kitawaki, "Perceptual qos assessment technologies for VoIP," *IEEE Commun. Mag.*, vol. 42, no. 7, pp. 28–34, Jul. 2004.
- [39] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton, NJ, USA: Princeton University Press, May 2004.
- [40] W. Wiegerinck and T. Heskes, "Probability assessment with maximum entropy in Bayesian networks," in *Proc. Comput. Sci. Statist.*, 2001.
- [41] W. Wu *et al.*, "Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework," in *Proc.* 17th ACM Conf. MM, Beijing, China, 2009, pp. 481–490.
- [42] I. Zukerman and D. W. Albrecht, "Predictive statistical models for user modeling," User Model. User-Adap., vol. 11, no. 1–2, pp. 5–18, 2001.



Arkady Zaslavsky is the Science Leader of the Semantic Data Management science area leading national and international projects in Internet of Things and sensor-based distributed systems. Before coming to CSIRO in July 2011, he held a position of a Chaired Professor in Pervasive and Mobile Computing at Luleå University of Technology, Sweden. Arkady received PhD in Computer Science from Moscow Institute for Control Sciences (IPU-IAT), USSR Academy of Sciences in 1987. Arkady has published more

than 300 research publications throughout his professional career and supervised to completion more than 30 PhD students.



Karan Mitra received the dual-badge Ph.D. degree from Monash University, Melbourne, VIC, Australia, and Luleå University of Technology, Luleå, Sweden, in 2013. He received the Minor Thesis (MIT) from Monash University, in 2006 and 2008, respectively. He received the BIS (Hons.) degree from Guru Gobind Singh Indraprastha University, New Delhi, India, in 2004. His current research interests include QoE measurement and prediction, context-aware computing, cloud computing, and mobile

and pervasive computing systems.



Christer Åhlund is a chaired professor with pervasive and mobile computing at Luleå University of Technology, Luleå, Sweden. He is also a scientific director of excellence in research and innovation named Enabling ICT. Beyond his academic background, he has 12 years of industry experience in the ICT area.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.