On the Relation between Customer Satisfaction and Performance Measurements

Tim Irnich, Tobias M. Kaufmann
Communication Networks, RWTH Aachen University
Kopernikusstr. 16, D-52074 Aachen, Germany
e-mail: tim|tmk@comnets.rwth-aachen.de

Volker Verkamp
P3 Solutions GmbH
Dennewartstr. 25–27, D-52068 Aachen, Germany
e-mail: volker.verkamp@p3-solutions.de

***Abstract*— This paper addresses the question if evaluation of service quality in wireless networks by means of *measuring* certain figures of merit allows to reliably derive the grade of satisfaction of mobile users. Core innovation of our approach is a questionnaire-based evaluation of the customer satisfaction immediately after each call. The questionnaire is displayed on the screen of the mobile phone. Evaluation of correlation between the customer satisfaction and *Measurement Report* (MR) parameters RXLEV and RXQUAL [1] recorded during each call gives new insights into the interrelation between objective figures of merit and customer satisfaction. The key result is that there is a weak relation between the user's speech quality perception and MR parameter values, but owing to the fact that for the whole range of MR parameter values RXLEV and RXQUAL the variance of speech quality ratings given by the users is high, is is not possible to predict the customer satisfaction by only considering technical performance measures. The conclusion is that a user-centric quality management for mobile services should be based on measurements of technical parameters and customer satisfaction instead of measurements of technical parameters only. The tool that has been used for the presented research has the potential of being used for this purpose.**

***Keywords*— Quality Monitoring, Speech Quality, User Satisfaction, SERVPERF, Correlation**

I. Introduction

Quality monitoring, evaluation and control of the mobile voice telephony service is currently performed by means of several kinds of statistical data obtained from network measurements, e.g. call blocking probabilities or dropped call rate. Speech quality is evaluated by *drive tests*, which are based on speech synthesis and speech recognition software (see e.g. [2]). This classical approach gives an *objective* picture of the quality of the voice telephony service. The service quality perceived by a customer is *subjective*. Target values for the measured parameters for objective evaluation are based on an assumed relationship between these parameters and customer satisfaction (see e.g. the *Satisfied User Criterion* [3]). A question that is left open by objective measurements is the actual grade of user satisfaction reached in a network that fulfills specific quality monitoring target values, or if the actual grade of user satisfaction can be predicted based on such results.

We address this question by simultaneously recording user satisfaction and network performance. This way for each service usage results representing both the technical and the user perspective are available and can be compared.

II. Investigation approach

In the scope of the presented research customer satisfaction is determined by evaluating the user's answers to a questionnaire, which is implemented via a *SIM Application Toolkit* (SAT) applet. The SAT-questionnaire is displayed on the mobile phone's screen immediately after each call.

In customer satisfaction research the applied measurement approach is called SERVPERF [4]. During the call GSM *Measurement Report* (MR) parameters *Receive Level* (RXLEV) and *Receive Quality* (RXQUAL) are recorded by the SAT-applet. The sample values of RXLEV and RXQUAL and the answers to the questionnaire are automatically sent to a server via SMS after the questionnaire is completed.

A pre-questionnaire is used to obtain demographic data needed for characterization of the participant group. SAT-Questionnaire and Pre-Questionnaire are designed with respect to well-established customer satisfaction measurement techniques [5], [4].

The results of MR parameter measurements and SAT-questionnaire allow to determine the subjective service quality perceived by the customer, and to compare it with the actually measured technical quality indicators (see Fig. 1).

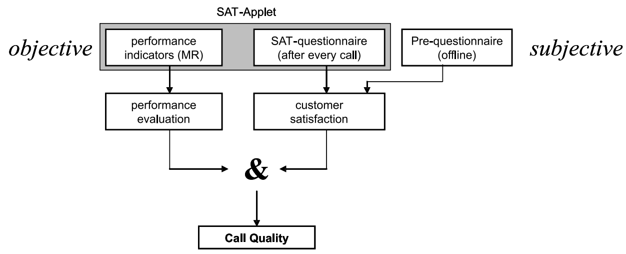


Fig. 1. Investigation approach

A. Objective performance measures

Objective performance indicators are considered in terms of GSM MR parameters RXLEV and RXQUAL [1]. RXLEV is the absolute field strength received by the mobile phone, while RXQUAL represents the technical quality of the radio link. In the GSM MR RXLEV values are coded in 64 discrete steps from 0 to 63, where 63 represents the highest electrical field strength. RXQUAL is coded in 8 discrete steps from 0 to 7, where 7 represents a high Bit error rate.

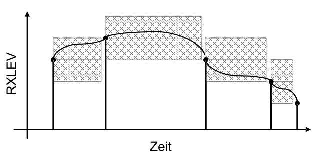


Fig. 2. Threshold based sampling of RXLEV during call

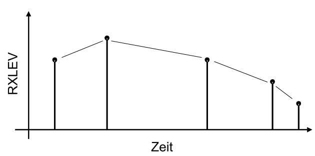


Fig. 3. Linear approximation of RXLEV based on sampling points

Owing to the limited processing power and memory provided by the SIM's processor for execution of the SAT applet, a threshold-based method for reducing the amount of sample values per call was used. This method is visualized in Fig. 2. Starting with the first RXLEV value, a new RXLEV value is stored every time the RXLEV deviates from the previous value more than a certain threshold. Linear approximation of the RXLEV behavior leads to a sufficiently accurate representation of RXLEV over the duration of the call.

Owing to the lower number of possible values, this kind of data reduction is not necessary for the RXQUAL parameter.

B. Subjective call quality perceived by the user

The user's speech quality perception is evaluated using a questionnaire that is displayed on the mobile phone's screen after every call attempt. In Fig. 4 a simplified flowchart of the questionnaire is shown. The circles in Fig. 4 indicate the position of additional questions not visualized here, because they are not relevant for the results presented in this paper. Based on signalling of events call setup, call connect and call release, and answers to previous questions the SAT applet selects one of several available versions of the questionnaire. This way the questionnaire can be adopted to suit the events that happened during the call.

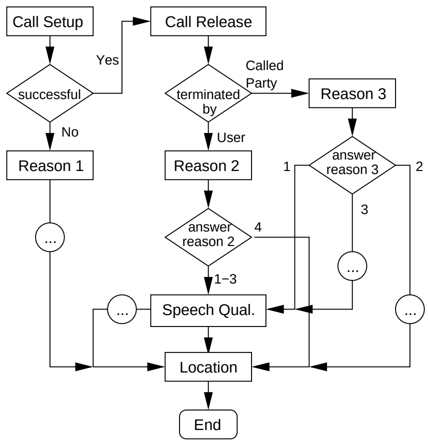


Fig. 4. Simplifi ed flowchart of the SAT-questionnaire

In case of an unsuccessful call setup attempt the user is asked for the reasons. Possible reasons are that the user voluntarily terminates the call setup ("wrong number dialed" and "terminated before partner picked up"), and occasions where technical problems prevent successful call setup ("I terminated" and "network terminated"); see Fig. 5(a). If the call is successfully established - the ring-tone is counted as already established call - after call release the SAT-questionaire distinguishes whether the user terminated the call, or if the network or the remote party terminated. In both cases the user is asked to specify if the call was released owing to technical problems; see Fig. 5(b) and 5(c).

In all cases where it can be assumed that some conversation took place during the call, the user is asked to rate the speech quality. The possible spectrum ranges from "excellent"to "bad"; see Fig. 6. In any case the user is asked to specify the type of location where the call took place; see Fig. 7.

"**SPEECH QUALITY 1**"
How do you evaluate the speech quality?
5. excellent
4. good
3. fair
2. poor
1. bad

Fig. 6. Speech quality rating

"**LOCATION**"
Where have you been?
1. Home Indoor
2. Office Indoor
3. Public Indoor
4. City Outdoor
5. Rural Outdoor
6. Car
7. Local Public Transport
8. Railway, Train (DB)
9. Elevator
10. Underground, Tunnel
11. Others

Fig. 7. Locations

III. Results

The survey was conducted for 60 days. During this time data was collected for 5717 calls, which is equivalent to 3.4 calls per person per day. For 2346 calls answers to the SAT-questionnaire were collected, for 1849 calls the SAT-questionnaire was completely answered. This results in an overall rate of return of 41 %.

A. Participant group

From the answers to the pre-questionnaire several statistics characterizing the group of participants in the survey have been obtained. 75% of the participants are male, 25 % are female. The average age is 37 years, the youngest participant is 29 and the oldest one is 55. Most of the participants (85 %) hold a university degree. 60 % of the participants stated that their predominantly use their mobile phone for business purposes, while 40 % of the participants have predominant private usage for private purposes, see Fig 8. With an average of 8.5 years of experience in mobile phone usage the participant group is very experienced as mobile customers. The least experience a participant specified was 4 years, and the most experienced participant has used mobile communication for 20 years; see Fig. 9.

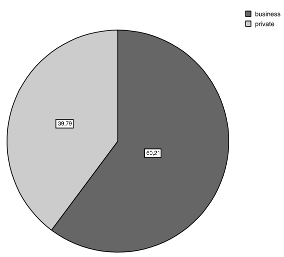


Fig. 8. Mobile phone usage

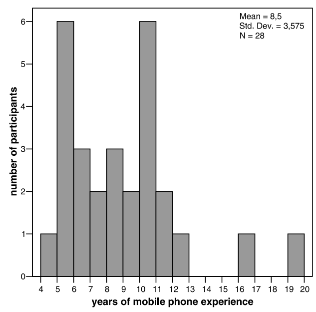


Fig. 9. Mobile phone experience

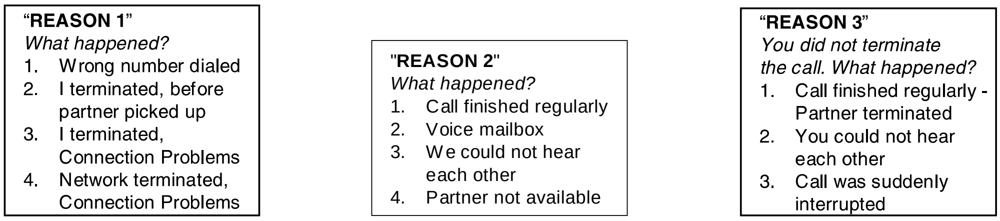
It is immediately clear that this participant group is not representative for all customers of a mobile radio operator. The group structure was intentionally selected, because the survey was targeted towards a proof-of-concept evaluation of the proposed methodology. As already stated above, the goal was to determine if this new approach provides new insights into the relation between customer satisfaction and call quality in terms of technical parameters. By selecting experienced participants with a strong technical background the probability of mis-interpreting effects observed during the call, and in consequence of giving inappropriate answers to the questions can be minimized. Furthermore the interaction with the SAT-applet can be considered to run smoother for this group than on average, leading to a higher rate of return.

B. Incidents of impaired call quality

Incidents of impaired quality that can be detected via the answers to the SAT-questionnaire are categorized as follows:

- **NoAccess** is an incident where the user initiates a call attempt, but the call is not established. Four different reasons for such incidents are distinguished in the SAT-questionnaire; see Fig. 5(a).
- **BadCall** is any case of an impaired call or call attempt that matches one or more of the following criteria:
  - Speech Quality evaluation *fair*, *poor* or *bad*.
  - NoAccess criterion fulfilled.
  - User terminates the call due to reason "we could not hear each other"; see Fig. 5(b).
  - Call is terminated by network or remote party, and answer to question "reason 3" is either "You could not hear each other" or "Call was suddenly interrupted"; see Fig. 5(c).

In Fig. 10 the relative frequency of occurrence of the different NoAccess incidents is visualized. 4.12 % of the calls have not been successfully established owing to technical reasons. Compared with the usual dimensioning target of 2 % call blocking rate, we can conclude that from the user perspective non-successful call attempts happen significantly more often than planned. Since



(a) Reasons for user-terminated call setup

(b) Reasons for user-terminated call

(c) Reasons for call terminated by remote party

Fig. 5. SAT-questionnaire questions for reporting of technical problems

for the connection attempts that were terminated by the user also technical reasons could have prevented the call being successfully established, this figure can be considered as a lower bound estimate for the actual NoAccess-rate.

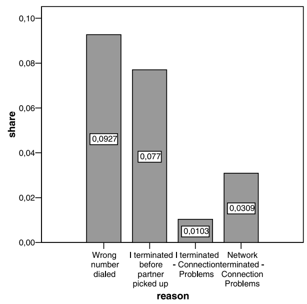
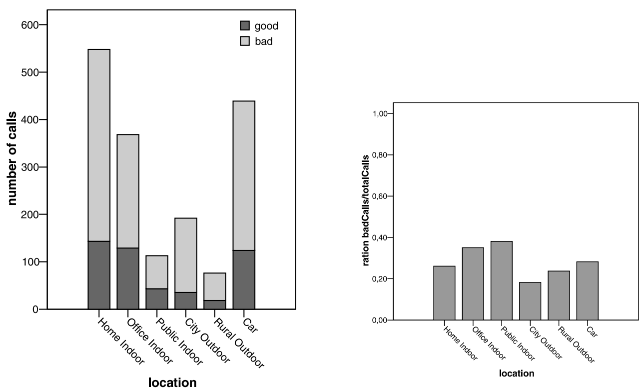


Fig. 10. Occurrence ratios of reasons for NoAcess incidents

Fig. 11 shows the statistics of the calls matching the BadCall criterion separately for the specified locations. Fig. 11(a) shows absolute numbers of BadCalls and total calls per location, while Fig. 11(b) displays the ratio of BadCalls to total calls per location. Locations with a low total number of calls are not shown due to low statistical significance.



(a) Fraction of BadCalls

(b) BadCalls / total Calls

Fig. 11. BadCall incidents

Dominant call locations are "home indoor", "office indoor" and "car". Indoor locations lead to a BadCall probability of approximately 25-40%, while in outdoor locations the range of BadCall percentage is approximately 18-30%.

Indoors public and office buildings have higher BadCall rates than private indoor. Possible reasons are different materials used for public and office buildings (e.g. reinforced concrete, metal-coated glass, etc.), and residential buildings, respectively, and size of buildings leading to a higher mean distance to the nearest outer wall of the building. Obviously achieving reliable radio coverage in large building structures is more difficult than in residential buildings.

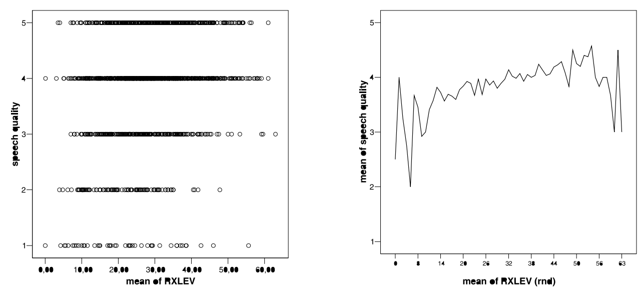
Not surprisingly it is visible that indoor locations in general have higher BadCall-rates than outdoor locations.

Radio coverage in dense urban areas appears to be more reliable than in rural areas.

Mobility (i.e. the "car" location) seems to have a significant impact. This location is somewhat special, because it is not possible to distinguish if the high BadCall rate is caused by attenuation (i.e. the car body), which would make this location comparable to the indoor cases, or if it is caused by the mobility, which would require the car location being compared with the outdoor cases. Assuming that the latter is true, mobility causes approximately 10% higher BadCall rate than the stationary cases.

C. Correlation of objective and subjective evaluation

In Fig. 12 the user's speech quality rating is shown over the mean RXLEV measured during the call. In Fig. 12(a) each circle represents one call, while Fig. 12(b) shows the mean of all calls that have the same mean RXLEV value on the y-axis. Obviously there is a weak linear relation between speech quality perception and RXLEV (see Fig. 12(b)), but the variance of speech quality ratings for calls that have the same mean RXLEV is considerable (see Fig. 12(a)). An RXLEV value between 15 and 46 leads with 95% probability and accuracy of ± 0.5 to a perceived speech quality between 3.5 and 4.2, which is equivalent to "good" speech quality. The correlation coefficient between mean speech quality rating and mean RXLEV per call is 0.225 with a high significance, which can be considered as a low correlation. From these results it is visible



(a) Scatter plot

(b) Mean speech qual. rating

Fig. 12. Speech quality vs mean RXLEV measured during call

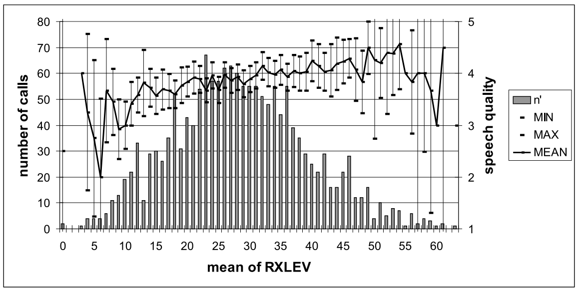


Fig. 13. Speech qual. vs. RXLEV, confidence intervals, value histogram

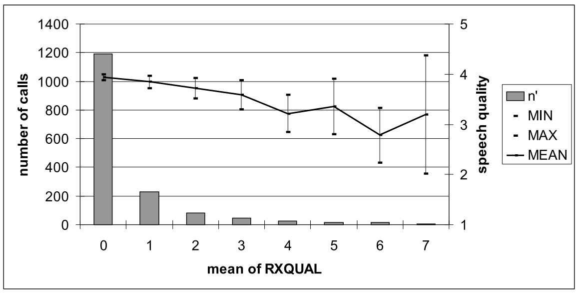
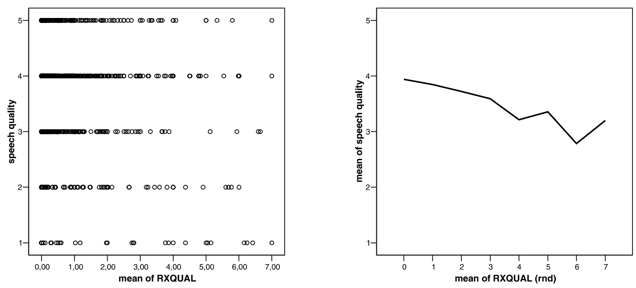


Fig. 15. Speech qual. vs. RXQUAL, confidence intervals, value histogram

that is not possible to reliably predict the speech quality perceived by the user in a differentiated way based on RXLEV measurements. Nevertheless the results indicate that it is possible to define limits for RXLEV that quite reliably lead to acceptable speech quality.

The ranges of very low and very high values for the mean RXLEV were expected to deliver most interesting results, because in these ranges impaired speech quality should be securely detectable by the participants. High RXLEV values cause impaired speech quality due to over-driven RF frontend. Unfortunately owing to low total number of occurrence the results in these ranges are of low statistical relevance (see Fig. 13). The low frequency of occurrence of these incidents can be explained by the fact that most calls took place in urban or dense urban areas (see Fig. 11(a)), where micro-cellular deployment is dominant.

In Fig. 14 the user's speech quality rating is shown over the mean RXQUAL measured during the call. Again in Fig. 14(a) each circle represents one call, while Fig. 14(b) shows the mean of all calls that have the same mean RXQUAL value on the y-axis. For the



(a) Scatter plot

(b) Mean speech qual. rating

Fig. 14. Speech quality vs mean RXQUAL measured during call

relation between perceived speech quality and RXQUAL a high variance of speech quality ratings is observed as well. The relation between mean speech quality and mean RXQUAL per call can be considered as a weak linear relation with negative slope.

The density of values decreases significantly with increasing RXQUAL values, which is also clearly visible from the width of the confidence intervals (probability 95%) in Fig. 15. For RXQUAL values larger than 4 it is no longer possible to derive reliable conclusions from the results. For RXQUAL value less or equall than 4 again 95% of all calls have a mean speech quality rating of $3.5 - 4.2 \pm 0.5$.

Equal to the comparison of RXLEV and speech quality rating, the results indicate that owing to the variance of speech quality ratings for calls that have the same mean RXQUAL it is not possible to reliably predict the user satisfaction based on RXQUAL measurements in an individual case, but based on such results target values for the upper limit of RXQUAL can be determined, which have the advantage of being verified with a direct evaluation by the users.

IV. Conclusions

The call quality in terms of BadCall and NoAccess event ratio (see Sec. III-B) significantly differs between indoor and outdoor locations. Within the indoor locations radio coverage in public areas and office buildings is worse than in residential buildings.

Furthermore user mobility appears to be a factor of significant impact. A possible conclusion could be that indoor coverage and improved coverage under high velocity conditions could be a rewarding field of future investments, but further investigations are needed before such conclusions can be made.

The comparison of speech quality rating and mean of MR parameters RXLEV (received level of field strength) and RXQUAL (bit error rate) per call leads to the overall conclusion that the customer satisfaction cannot be predicted from parameters measurements in an accurate and differentiated manner. However, the applied investigation approach has shown the potential to be useful for determination and assessment of suitable target values for performance indicators that provide the basis for network operator quality monitoring.

The biggest advantage over existing approaches is the consistence of time and events objective and subjective parameters are based on. Any other approach for quality monitoring requires assumptions on the relation between the test conditions and the real-life-conditions the user is facing.

Future applications of the presented approach could serve multiple purposes. Large scale application could lead to a significantly more representative picture of overall service performance and quality, and to an improved understanding of the relation between technical service performance and customer satisfaction. Furthermore, the presented technique could be used as a "magnifying-glass" in case specific problems in the mobile network are reported or suspected at a given location and/or time. This could be particularly useful in the context of customer complaint management.

Technical limitations of the current SAT-based measurement applet are expected to be overcome with the availability of measurement client software that is based on the Symbian mobile phone operating system.

The SERVPERF approach used to obtain customer satisfaction values delivers suitable quantities for comparison with objective parameter values. Also known from customer satisfaction research are the more advanced SERVIMPERF and *Customer Satisfaction Index* (CSI) approach. Their application requires a multi-dimensional evaluation of the speech quality. This could be done via splitting up the speech quality in analogy to the quality evaluation technique *Readability, Signal Strength and Tone Quality* (RST), which is applied in amateur radio communication.

Additional insights should be possible when more samples are available, especially in the range of extremely high or low values of RXLEV and RXQUAL. Since these values are rare, a longer evaluation period and a significantly higher number of users is required for future surveys.

In future surveys more attention should be spent on the spatial distribution of participants. During the survey the users were predominantly located in urban or dense urban areas, where the mobile radio network is usually well-engineered. A representative picture of the whole network can only be obtained if also users that work and live in rural areas take part in the survey.

ACKNOWLEDGEMENTS

The authors would like to thank their partners at Ericsson GmbH Dusseldorf for the good cooperation and fruitful discussions.

REFERENCES

[1] ETSI/3GPP, "Radio Access Network; Radio Subsystem Link Control, V6.6.0, Release 6," Technical Specification 3GPP 45.008, Feb 2004.

[2] ITU-T, "Perceptual evaluation of speech quality (PESQ)," ITU-T Recommendation P.862, Feb 2001.

[3] ETSI/3GPP, "TR 101 112, Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS (UMTS 30.03)," Tech. Rep., Apr. 1998.

[4] A. Parasuraman, V. Zeithaml, and L. Berry, "A conceptual model of service quality and its implications for future research," *Journal of Marketing*, vol. 49, pp. 41–50, 1985.

[5] J. Cronin and S. Taylor, "Measuring service quality: A reexamination and extension," *Journal of Marketing*, vol. 56, pp. 55–68, jul 1992.