Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2015) 23: 548 – 559
© TÜBİTAK
doi:10.3906/elk-1302-191

Quality of service assessment: a case study on performance benchmarking of cellular network operators in Turkey

Rana KADIOĞLU[1], **Yaser DALVEREN**[2], **Ali KARA**[2,*]
[1]Roll-Out Management-Central Anatolia, Vodafone, Ankara, Turkey
[2]Department of Electrical and Electronics Engineering, Faculty of Engineering, Atılım University, Ankara, Turkey

Received: 26.02.2013 • **Accepted:** 17.04.2013 • **Published Online:** 23.02.2015 • **Printed:** 20.03.2015

Abstract: This paper presents findings on performance benchmarking of cellular network operators in Turkey. Benchmarking is based on measurements of standard key performance indicators (KPIs) in one of the metropolitan cities of Turkey, Ankara. Performance benchmarking is formulated by incorporating customer perception by conducting surveys on how important KPIs are from the user's point of view. KPIs are measured, with standard test equipment, by drive test method on specified routes. According to the performance benchmarking results, the GSM and UMTS network operators achieving the best performance were determined in Ankara. Speech qualities of network operators, as the most popular service, were also evaluated by several statistical methods including pdf/cdf analysis and chi-square and Fisher's exact tests. The network operator providing the highest speech quality in Ankara was determined with the methods applied. Overall, the results and approaches on benchmarking of cellular networks in Turkey are reported for the first time in this paper. The approaches proposed in this paper could be adapted to wide-scale benchmarking of services in cellular networks.

Key words: Cellular networks, benchmarking, key performance indicators, speech assessment, quality of service

1. Introduction

Customer satisfaction is critical to gain a sustainable competitive edge in the market. In communication networks, as the customer's satisfaction with the service is directly dependent on the quality and the performance of the network, measurements of network performance and quality of service (QoS) assessments are crucial [1–7]. For this purpose, network operators should survey the performance of their networks and measure quality parameters on a regular basis as customers' needs and satisfaction are presumably the main market driver, and especially in wide-area service networks such as cellular communications networks [8]. Generally, network optimization engineers exert efforts to increase the quality and capacity of operational networks, and to develop and deploy new services in order to meet customer demands and to guarantee customer satisfaction. Key performance indicators (KPIs) are universally accepted parameters of cellular networks that engineers need to survey and keep within some specified threshold values in order to meet the QoS criteria required by both competent authorities and customers [8–10]. On the other hand, benchmarking is a comparison of a company's network performance with their competitors' and/or a comparison of its performances measured at different periods in time. In other words, benchmarking is a kind of assessment of the service quality of an operational network [11].

[*] Correspondence: akara@atilim.edu.tr

The performance of network and service quality is frequently assessed with the drive test method [12,13]. This method uses measurement results and operational information about a network with drive tests on some routes, and the outcomes are used to assess the network performance for possible improvements to the network. For this purpose, network operators periodically perform such assessments in a specified geographical area to assess their own performance as well as to compare it with their competitors. This is a kind of QoS assessment, and its results could also be used for optimization, development, and possible service extensions to the network.

This paper presents results of a benchmarking study of cellular network operators in Ankara, Turkey. The study provides the following contributions: 1) customer satisfaction and/or needs are integrated into the benchmarking process and thus the QoS assessment in order to see how customers perceive QoS; and 2) an improved approach is given for speech quality evaluation of cellular network operators.

The remainder of the paper is as follows: the second section introduces the concept of QoS in cellular networks, and relevant parameters and factors along with KPIs used in the measurements are described. Section 3 presents the measurement methodology, including selection of routes, measurement systems and procedures, etc. In Section 4, field measurements are described, and derivations of KPI values along with a highly efficient mathematical approach based on the use of standard KPIs are presented. Performances of cellular network operators are then compared on the basis of the proposed approach. Section 5 discusses statistical techniques for speech quality evaluation and speech benchmarking of cellular network operators. Finally, conclusions are drawn in the last section.

2. QoS assessment in cellular networks

QoS can be described as the ability of a network to provide a service at an assured service level. This can be measured by either the network operator itself or by an independent or regulatory organization. QoS is very critical in cellular communication networks, including second-generation (2G) and third-generation (3G) systems. As 2G networks were unable to provide better and faster data services, 3G system have been deployed to provide a variety of data services such as internet browsing, e-mails, video telephony, and video streaming (dense data needs such as YouTube or Instagram). Currently, both 2G and 3G cellular networks have become operational in most countries. Almost all current operational cellular networks support both circuit-switched (CS) and packet-switched (PS) services, and QoS assessment of CS and PS services should be evaluated separately [8]. Customer surveys and feedback may be used in assessment of QoS as perceived by customers [3,11]. As the objective of this paper, QoS is evaluated by customer perception via customer-oriented benchmarking tests.

QoS assessment is based on QoS parameters, which can be given with a layered structure [11]: 1) the first layer represents network availability as the QoS from the service provider's point of view; 2) the second layer represents network access as the basic requirement from the user's point of view; 3) the third layer represents various QoS aspects, including service access, service integrity, and service retainability; 4) the fourth layer represents different services whose outcomes are the QoS parameters; and, finally, 5) the last layer represents KPIs of each service of the fourth layer.

As shown in [14,15], KPIs constitute the bottom layer, and benchmarking is mostly conducted with standard KPIs that have been used for many years. In drive tests, KPIs were obtained for each network operator by measurements performed on specified routes. Some of the KPIs and their descriptions are, as they are used in benchmarking, listed in Table 1.

Table 1. KPIs and their descriptions for CS and PS services.

Service	KPI	Description
CS (Voice)	Call setup success rate (CSSR) (%)	The ratio of successful call setups to call attempts
	Call setup time (CST) (s)	Duration to completing address information
	Dropped call rate (DCR) (%)	The ratio of number of dropped calls to successful calls
	Speech quality (PESQ)	Quality of speech as perceived by users
	Received signal level (RxL) (dBm)	Received signal power at the input of the mobile device
PS (Data)	Attach success rate (ASR) (%)	Probability that a subscriber can attach to the network
	Attach time setup (ATS) (s)	Time duration taken to attach to the network
	Packet data protocol (PDP) context activation success rate (CASR)	Probability that subscriber can activate a PDP context (%)
	PDP context service access success rate (SACR) (%)	Probability that a subscriber access the service successfully
	Service session success rate (SSSR) (%)	Probability of initiating the service by the subscriber
	FTP data throughput (FTP DL) (kbps)	Average data rate that can be achieved

3. Measurement methodology

In the measurements, the standard drive test method was used in data collecting, from which performance of the network and service qualities were assessed. Drive tests were conducted with a mobile vehicle containing measurement equipment to measure relevant indicators related to QoS of a cellular radio network in a given geographical area. There was one measurement device for each network operator in the mobile vehicle and two subscriber identity module (SIM) cards for each equipment set. One of the SIM cards was used for voice services (SC) and another was used for data services (PS). Moreover, a scanner was connected to the system to scan relevant bands for detecting active frequency channels along the route. A global positioning system (GPS) was also connected to the system to get geographical positional information. Generic components of a drive test system for QoS measurements can be found in [13,15] in detail, and a typical diagram is shown in Figure 1.

In a typical drive test procedure, a test vehicle travels on a predetermined route. In these measurements, the drive test routes were determined in the city of Ankara, as shown in Figure 2, and its suburbs where various range of user profiles could be sampled.

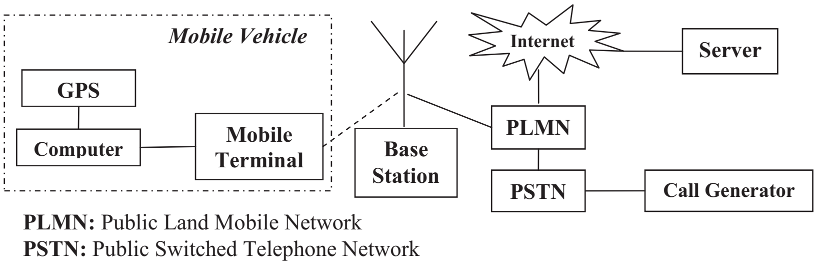

Figure 1. Typical Block diagram for drive test system.

All measurements were started at 0800 hours and ended at 1800 hours. In such measurements, the choice of routes is very important for the measurement quality as it affects both the results and the analysis to be performed later. For this purpose, selected routes in the measurements typically contained a highway, a highway entrance and exit, a highway connection road, a main boulevard, a trade center, and high user clarity areas in the city. Hot spots, dense urban areas, and urban areas were also included in the routes. Moreover, the routes were determined to have minimum stops for completing measurements in a continual way. Alternatives and bypass routes were then determined in advance in case of road constructions, traffic jams and/or accidents, etc.

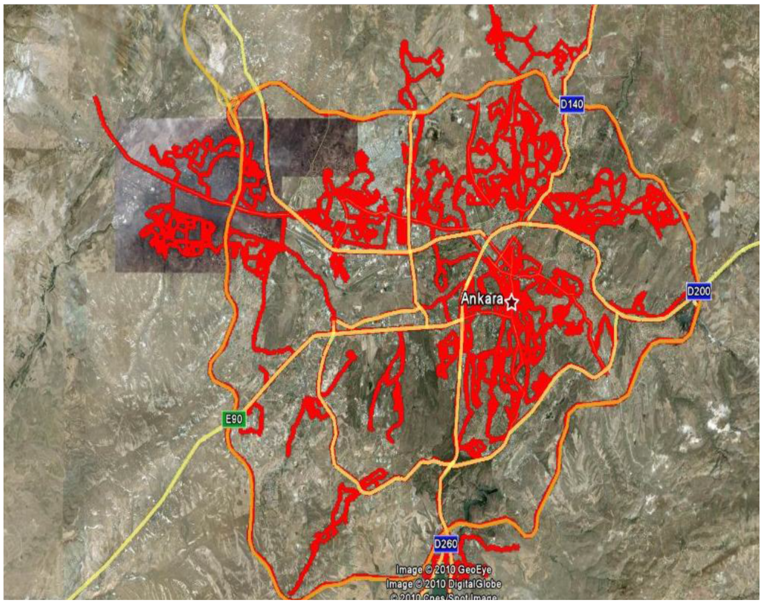

Figure 2. Drive test route in Ankara.

In such field tests, all measurements are made automatically. In other words, preprogrammed work orders showing all calls, idle durations, and cycles during the measurement are loaded to the measurement equipment from the server. A work order is also prepared for the scanner as it determines the frequency channels over which the calls are made. In the field tests, typically, a 90-s call is made, and then it is followed by a 20-s idle, and so on. As the measurement should be a continual process, this is repeated along the route. All logs regarding the measurement parameters are then automatically sent to the server in order to be analyzed later. Measurements can be split up into two categories for description: voice (CS) and data (PS).

For a typical voice measurement, call generators, like that in Figure 1, place automated calls to the cellular network operator and cellular terminal. A cellular terminal means that a call is placed from the call generator to a mobile device, whereas a cellular network operator means that a call is placed from a mobile device to the call generator. As soon as all recorded logs are passed to the server, the data to be processed by server are accessible by the computer. In this way, all KPIs described in the previous section can be evaluated. However, among the KPIs, only Rx level data is obtained by the logs recorded on the scanners.

For a typical data measurement, a data card is used for each network operator. A preprepared 2-MB upload file and 5-MB download file are transmitted to, and then are received from, the FTP server. For this purpose, first, the PDP is activated to provide communication between the mobile device and network server. After PDP context activation, a 5-MB test file is received from a public server, as it is connected to the internet, and then a 2-MB test file is sent to the public server (FTP upload). HTTP-based web browsing is generally used for testing. A 1-MB website is determined before web browsing, and then 10 ping commands with 32-Kbyte packet size are used to make round-trip delay measurements statistically. Finally, the PDP context is deactivated while there no data activity over the packet protocol context.

4. Measurement-based QoS assessment

This section provides a methodology that can be used to evaluate QoS and benchmarking of cellular network operators in Turkey. As discussed in previous sections, end user perception of service quality is integrated into QoS assessment. It is known that KPIs described in the previous section are used to judge network performance and evaluate QoS. As discussed in Section 2, the upper-level indicators for QoS performance are accessibility and retainability, as they form the upper-level layer of ETSI QoS structure. Both are related to the KPIs described in the previous section. Accessibility is the probability that a service can be obtained within given conditions when requested by a service user. Accessibility is measured by number of call attempts (#CA) and number of successful calls (#SC) in the tests, and it can be formulated as:

$$\text{accessibility} = \frac{\#\text{SC}}{\#\text{CA}}. \tag{1}$$

On the other hand, retainability is described as the probability that an active call terminates successfully. It is estimated as the ratio between the number of successful calls (#SC) and number of normally terminated calls (#NTC). It can be expressed as:

$$\text{retainability} = \frac{\#\text{NTC}}{\#\text{SC}}. \tag{2}$$

Using Eqs. (1) and (2), some of the KPIs can be calculated from the measured data. From the definition of CSSR in Table 1, the calculated value of accessibility according to Eq. (1) can be used to represent CSSR. DCR from the same table can be formulated as retainability of the network using the following relationship:

$$DCR = 1 - \ \text{retainability}. \tag{3}$$

In order to obtain accessibility, retainability, and DCRs for cellular network operators, test call samples presented in Table 2 are used. Some of KPIs from Table 1 are then obtained as in Table 3, where directly measured KPIs for each network are also presented.

Table 2. Test call samples obtained from drive tests.

Operator	# CA	# SC	# DC	# NTC
A	177	173	1	172
B	173	172	2	170
C	193	189	1	187

Table 3. Comparison of CS KPIs (GSM & UMTS/3G network).

CS KPIs	GSM network			UMTS/3G network		
	Operator A	Operator B	Operator C	Operator A	Operator B	Operator C
CSSR (%)	97.74	99.42	97.92	98.86	100	100
CST (s)	2.50	1.87	2.59	2.97	2.17	2.64
DCR (%)	0.57	1.16	0.52	0.58	0.60	0.00
PESQ (avg.)	3.48	3.66	3.83	3.93	3.71	3.84
RxL (dBm)	–76.5	–73.5	–73.9	–97.5	–98.8	–99.0

PS network KPIs are also listed in Table 4. These results are based on standard methods proposed in the literature.

Table 4. Drive test results of PS services.

KPIs	Operator A	Operator B	Operator C
ASR (%)	100	100	97.8
AST (%)	2.53	1.62	2.02
CASR (%)	100	100	100
SACR (%)	89.4	95.0	96.7
SSR (%)	90.32	96	100
FTP DL (Mbps)	0.11	0.10	0.10

It should be noted that end user or customer perceptions of KPIs and QoS are not taken into account in these standard assessments. Therefore, these figures could be used to determine “equivalent weighted scores” of networks or operators, as customer perception is not used to weigh each KPI. In order to accompany how customers perceive the QoS of a service, customer surveys could be conducted to collect customer experiences and satisfaction levels of the performance of cellular network operators [3,6–8]. For this purpose, a simple questionnaire with a list of service quality indicators was given to customers. In essence, each service quality indicator corresponds to one KPI listed previously. Service quality indicators were then graded by customers. For example, DCR, as a KPI, was translated to “dropping a call” in the questionnaire, while RxL (dBm) was, from the customer’s point of view, “coverage” or “out of service” for the network. Surveys were conducted in areas where density of customers and service usage were high (for example, Kızılay and universities). The question was simple: “Can you assign points to the following indicators according to the significance levels from your point of view in preferring a certain operator?”

The survey was completed by 800 respondents, and each gave points between 0 and 100 to each indicator (or quality level of the network operator). After analysis of the data, priority of the indicators was obtained as in Table 5. These figures can then be incorporated into the performance assessment of cellular networks. In this way, a weighted performance score for each network can be calculated along with equivalent weighted performance (without customer weighting of KPIs). To obtain the weighted scores, the method is straightforward: weighted scores are computed by the product of field test values and figures from customer survey results in Tables 3 and 5, respectively. Simple formulas used in calculations of weighted scores of KPIs are provided in Table 6.

Table 5. Ratings of customers of KPIs (CS).

CS KPIs	Percentage (%)
CSSR	23.9
CST	2.7
DCR	15.9
PESQ	18.6
RxL	38.9

In calculating the weighted scores, the following should be considered:

- Lower CST indicates higher performance. In order to take this negative correlation into account, the test value is reversed.
- Minimum call setup value is assumed to be 1 s in CST calculations of Table 6.
- Reference received signal level (RxL) in Table 6 is taken as –120 (in dBm). It corresponds to the maximum Rx sensitivity of 2G. This value should be changed for 3G (–136 dBm).

- Moreover, the A-PESQ value is assumed to be 4.5 in the PESQ calculations of Table 6. However, a detailed discussion on PESQ is provided for speech evaluation in the next section.

  [-] Note that the sum of the weighted scores of KPIs becomes 100, which represents the highest quality according to this formulation. Therefore, the best score would be a figure close to 100.

Table 6. Expressions to determine KPIs in terms of weighted scores.

KPIs	Expressions
CSSR	$\text{CSSR} = \left(\frac{\text{T-CSSR}}{100}\right) \cdot \text{S-CSSR}$
CST	$\text{CST} = \left(\frac{1}{\text{T-CST}}\right) \cdot \text{S-CST}$
DCR	$\text{DCR} = \left[1 - \left(\frac{\text{T-DCR}}{100}\right)\right] \cdot \text{S-DCR}$
PESQ	$\text{PESQ} = \left(\frac{\text{T-PESQ}}{\text{A-PESQ}}\right) \cdot \text{S-PESQ}$
RxL	$\text{RxL} = \left(\frac{\text{T-RxL} + 120}{100}\right) \cdot \text{S-RxL}$

The first letter of the KPIs, "T-" or "S-", represents the source of the KPI, whether it is a test result (T) or customer survey result (S). "A-" represents an average value.

With the help of Table 6, weighted KPI values and rankings for GSM and UMTS of network operators are calculated. These are shown for GSM and UMTS networks in Table 7.

Table 7. Weighted KPI values and ranking (UMTS network).

CS KPIs	GSM network			UMTS network		
	Operator A	Operator B	Operator C	Operator A	Operator B	Operator C
CSSR	23.36	23.76	23.4	23.63	23.9	23.9
CST	1.08	1.44	1.04	0.91	1.24	1.02
DCR	15.81	15.72	15.82	15.81	15.8	15.9
PESQ	15.83	15.13	14.38	16.24	15.33	15.87
RxL	16.92	18.09	17.93	11.9	7.44	7.4
SCORE	73.00	74.14	72.57	72.57	71.06	71.08
Ranking	2	1	3	1	3	2

As can be seen from the rankings of three network operators (A, B, and C), Operator B's GSM network gets the highest score, while Operator A's UMTS network gets the highest score among the three.

In addition to the calculated weighted scores, equivalent weighted scores for KPIs are calculated. Equivalent weighted score is calculated by assuming that users rate all KPIs equally. In other words, each given KPI in Table 5 has a weight of 20%. Equivalent scores are then calculated as 73.16, 75.92, and 71.88 for GSM networks of operators A, B, and C, respectively, while they are 75.74, 73.02, and 72.03 for UMTS networks of operators A, B, and C, respectively. It is interesting that there is a very little change in the ranking as a result of two different scoring approaches; the UMTS network of operator C performs slightly better than the UMTS network of B when weighted scoring is used. As the variation of the scores is very limited, this result should be expected. From the statistics of services used in GSM and UMTS networks, the most popular service, as expected, is voice communication, or voice telephony. This makes, then, speech quality very important from the customer's point of view. Therefore, speech assessment of networks has been studied separately. In order to evaluate and benchmark speech quality of the networks of operators A, B, and C, further analysis on measurements was conducted. The following section presents the results of this study.

5. Speech quality evaluation and benchmarking

Speech quality refers to the clarity of a speaker's voice as perceived by a listener. Quality, in general, is formulated as the difference between perceptions and expectations [12,14,16]. Similarly, speech quality can also be defined as the result of a subject's perception of a conversation. In this way, speech quality measurement is a mean of measuring customer experience of voice telephony services. At the meeting of ITU-T Study Group 12 in February 2001 [12], PESQ was officially approved as new International Telecommunication Union (ITU-T) recommendation P.862 for PESQ use in objective testing where an original reference file sent into the system is compared with the impaired signal that comes out. This testing method provides an automated test mechanism that does not rely on human interpretation for result calculations. PESQ measures the effect of end-to-end network conditions, including CODEC processing, jitter, and packet loss. Therefore, PESQ is the preferred method of testing voice quality in an IP network. On the other hand, a new algorithm (POLQA, "Perceptual Objective Listening Quality Analysis") was developed recently by leading industry experts for speech quality measurement [7,16]. POLQA is the next-generation voice quality testing technology for fixed, mobile, and IP-based networks. POLQA was standardized by the ITU-T as new recommendation P.863 and can be applied for voice quality analysis of HD voice, 3G, and 4G/LTE networks. However, PESQ was used in this work as the POLQA has not yet been adopted by operators for benchmarking processes.

In this regard, the contemporary measurement tool, PESQ, compares an original signal with the degraded version after sending it through a channel. It is able to predict speech quality with high correlation to subjectively perceived quality in terms of mean opinion score (MOS) in a wide range of conditions. PESQ is accepted as the industry standard for end-to-end speech quality measurements, and it is necessarily used to measure speech quality in cellular networks. The measurement system shown in Figure 1 could automatically generate PESQ values and record them for further analysis for the networks of operators A, B, and C.

There may be two assessment techniques based on the PESQ concept for a set of speech quality measurements, and these are used in the speech quality assessment of the networks in this study. One of the methods is statistical inference by means of probability density function (PDF) or cumulative distribution function (CDF). In this method, the PDF and the CDF of measured MOS scores are used to evaluate quality of speech, based on measurement results of speech samples. In other words, rankings of speech qualities of the networks are compared on the basis of how the MOS value of each network is distributed. For objective speech quality measurements, the MOS value here is calculated from speech samples. In the measurements, electronic maps through the drive test route are divided into bins of desired sizes. A bin is the minimal geographic unit of the electronic map, and the following typical bin sizes were employed: 20 × 20 m, 50 × 50 m, or 400 × 400 m. MOS values are given to all speech quality samples from each bin. An average MOS value is then calculated for each bin. Thus, PDF/CDF histograms of MOS values measured from the networks were calculated for quantitative assessments.

PDFs constructed from the collected data are shown in Figure 3. Corresponding CDFs of the MOS values of three network operators were obtained as in Figure 4.

It may be difficult to compare the networks from the PDFs. However, it can be concluded, in this case, that network operator B shows a better distribution, as MOS values are more concentrated between 3.5 and 4.0 and the maximum MOS value is slightly greater than those of the others. When the CDFs of MOS values of three network operators in Figure 4 are considered, it is evident that the speech quality of operator B is better than the others.

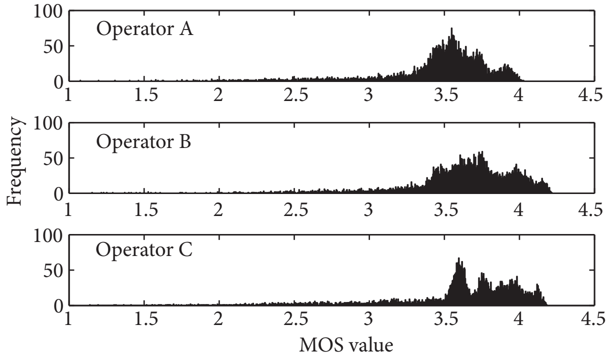

Figure 3. MOS PDFs of three network operators.

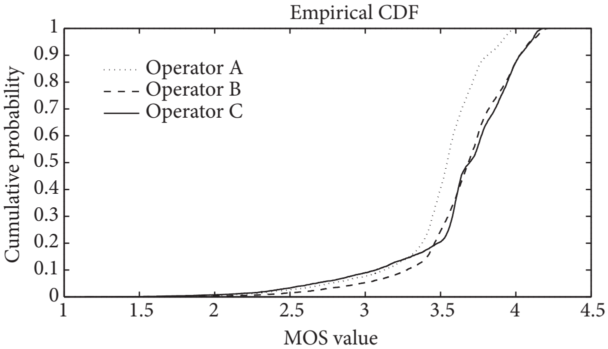

Figure 4. MOS CDF of three network operators.

In the second technique, hypothesis tests are considered to compare network operators' performances. For this purpose, two different hypothesis testing methods were used. These were based on classical statistical approaches, namely, chi-square (χ^2) and Fisher's exact tests. First contingency tables were obtained. Then, by using chi-square distribution, hypotheses tests were derived to compare network operators' performances. Finally, hypotheses tests were repeated using Fisher's exact test, which is a reliable method if the sample sizes are small.

Chi-square distribution is one of the most widely used probability distributions in inferential statistics, e.g., in hypothesis testing, or in construction of confidence intervals. The chi-square distribution is the distribution of a sum of the squares of k independent standard normal random variables. In many problems, it is not assumed that the available observations come from a particular parametric family of distributions. The chi-square test is widely used in cases where no special assumptions are made about the form of the distribution. In the chi-square tests conducted for speech comparisons of the networks, contingency tables were generated to examine dependency of the networks. It is known that a contingency table, as it is typically represented by a table having R rows and C columns (R × C), is a tabular arrangement of count data representing how much row factors are related to column factors. In this study, samples from the three networks are selected at randomly. Efforts focus on assessing paired observations (number of good speech samples/number of samples), and results are reported in a 2 × 2 contingency table that are independent of each other, as discussed in several works in the literature. In order to define the number of good samples, a threshold value is needed. This value indicates the minimum PESQ value that represents the acceptable speech quality. Thus, for a set of speech samples, each sample is compared with the threshold, and then good samples among the total number of samples can be obtained. Three different contingency tables were constructed since the tests were performed two by two for pairs of networks or operators. That is, operator A is compared with operator B and then operator C separately, and finally operator B is compared with operator C. Constructed contingency tables are given in Table 8.

Table 8. Contingency tables for operators.

	Operators A & B		Operators A & C		Operators B & C	
	A	B	A	C	B	C
Good samples (k)	7955	8892	7955	8247	8892	8247
Total samples (N)	8984	9653	8984	9430	9653	9430

On the basis of these contingency tables, hypothesis H_{XY} says that the performance of network X does not differ significantly and is not independent of network Y, while H'_{XY} says that the performance of network X differs significantly and is independent of network Y. In the case of 2 × 2 contingency, χ^2 is calculated from the following expression:

$$\chi^2 = \frac{(N_X + N_Y)(k_Y N_X - k_X N_Y)^2}{N_X N_Y (k_X + k_Y)(N_X + N_Y - k_X - k_Y)}, \tag{4}$$

where N denotes the total sample number and k denotes the good sample number for network operators X and Y. For instance, for the contingency table between network operators A and B, N_A represents the total sample size of network operator A (8984), while k_A represents the good sample size (7955). Thus, χ^2 tests are applied accordingly, and results are provided in Table 9.

Table 9. Test results for the networks.

Operators	χ^2
A & B	68.3
A & C	0.0021
B & C	113.29

Now, for testing the hypotheses, a threshold for the test statistic is determined as 3.841, where chi-square with 1 degree of freedom at 0.95 is assumed (significance level of 0.05). Then, for example, for networks X and Y, H_{XY} is rejected if the test statistic exceeds this threshold.

According to the test results in Table 9, H_{AB} and H_{BC} are rejected, whereas H_{AC} is accepted. Thus, it is concluded that the performance of network operator B is independent of the other two network operators. On the other hand, the performances of network operator A and network operator C are not significantly independent.

The χ^2 test is a reliable test in this research since the number of observations exceeds 5 in the above contingency tables. However, if the sample sizes are small, Fisher's exact test could be used as a statistical significance test in the analysis of contingency tables. Fisher's test is one of the exact tests as the significance of the deviation from a null hypothesis can be calculated exactly. The hypotheses tests were then repeated using Fisher's exact test. Accordingly, the probabilities, p, for the contingency tables are calculated from

$$p = \frac{\begin{pmatrix} N_X \\ k_X \end{pmatrix}\begin{pmatrix} N_Y \\ k_Y \end{pmatrix}}{\begin{pmatrix} N_X + N_Y \\ k_X + k_Y \end{pmatrix}}, \tag{5}$$

and the results are listed in Table 10.

Table 10. Probabilities for the contingency tables by using Fisher's test.

Operators	p
A & B	0.52
A & C	0.5
B & C	1

As can be seen from Table 10, calculated probabilities are not less than the significance level (0.05). Therefore, H_{AB}, H_{AC}, and H_{BC} are not rejected. Thus, it is concluded that the performances of networks A, B, and C are independent.

According to the χ^2 test, all networks can be compared in pairs, and if the performance differs significantly, they receive one point. For example, performances of the network operators A and B, as H_{AB} is rejected, both receive a “1” point. When the network operators A and C are compared, H_{AC} is not rejected, and then both receive a “0” point. The results are provided in Figure 5, where the networks are compared in pairs by hypotheses testing.

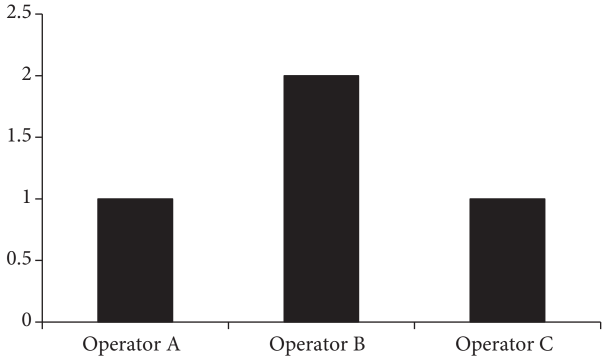

Figure 5. Pairwise ranking of networks.

6. Conclusions

There are three operators (networks) in Turkey, and standardized KPIs for the most popular services were used in the benchmarking process reported in this work. In the literature, typically, the service quality assessment does not include customer perception. Customer perception would require extensive analysis on how customers view the QoS given by a network operator. For this, one way would be to conduct surveys with the customers on the standard KPIs: how they grade the standard KPIs used in benchmarking. This approach would work well in incorporating customer perception into the benchmarking process. In this work, results of such a survey conducted with the customers in the city were used to weigh KPIs in assessment of the overall network quality. As a result, a customer weighted score and equivalent weighted score were obtained for ranking the networks. Although initial tests show that weighted scoring may not greatly affect the assessment results, there is still a need for incorporating customer perceptions of QoS in cellular networks. This study, then, could form a basis for future works as the method used brings a new approach to benchmarking. Wide-scale benchmarking of operators in a country would require extensive works, which could be administered by the regulatory organization of the country. For example, surveys could be extended to cover customers in different geographical regions, customer profiles, their expectations of services and requirements, etc.

When QoS is considered, speech quality is probably the most critical criterion in cellular networks, as speech still has the largest share among the services of cellular networks. This work therefore included evaluation of speech quality of network operators in the targeted city. For this purpose, different methods proposed in the literature were used to compare speech quality of networks. All methods for speech quality assessment are based on the use of MOS figures of speech samples. The most common method is to use the PDF and CDF of MOS scores. This method is simple but would be limited in giving an accurate assessment. There have been several statistical approaches for more accurate speech quality assessment. By utilizing these statistical methods, it is possible to determine independency of the operators and then compare them in pairwise rankings. The results obtained in benchmarking showed that network operator B achieved the highest performance of speech quality among the three network operators.

Finally, it should be noted that benchmarking should be conducted periodically, and currently operators prefer to monitor their own and competitors’ services by themselves and keep the results they obtain classified. In order to reduce customer complaints and confusion about services and to increase their satisfaction, regulatory organizations should take more responsibility in the assessment of networks and services given. This could be achieved by either planning regular benchmarking studies or subcontracting benchmarking works to third parties (some authorized companies with nondisclosure agreements), and then reporting some relevant outcomes and findings to the public.

References

[1] Awada A, Wegmann B, Viering I, Klein A. Optimizing the radio network parameters of the long term evolution system using Taguchi's method. IEEE T Veh Technol 2011; 60: 3825–3839.

[2] Kang K, Jeon WJ, Park KJ, Campbell RH, Nahrstedt K. Cross-layer quality assessment of scalable video services on mobile embedded systems. IEEE T Mobile Comput 2010; 9: 1478–1490.

[3] Carlos E, Otero IK, Luis DO, Scott LM. Characterization of user-perceived quality of service (QoS) in mobile devices using network pairwise comparisons. International Journal of Wireless & Mobile Networks 2010; 3: 141–153.

[4] Fraimis IG, Kotsopoulos SA. QoS-based proportional fair allocation algorithm for OFDMA wireless cellular systems. IEEE Commun Lett 2011; 15: 1091–1093.

[5] Fuzheng Y, Shuai W. Bitstream-based quality assessment for networked video: a review. IEEE Commun Mag 2012; 50: 203–209.

[6] Afullo TJO. Quality of service in telecommunications - the customer's perspective. In: IEEE 7th Africon Conference; 17 September 2004; Gaborone, Africa. New York, NY, USA: IEEE, 2004. pp. 101–106.

[7] Jelassi S, Rubino G, Melvin H, Youssef H, Pujolle G. Quality of experience of VoIP service: a survey of assessment approaches and open issues. IEE Communications Surveys & Tutorials 2012; 14: 491–513.

[8] Chevallier C, Brunner C, Garavaglia A, Murray KP, Baker KR. WCDMA (UMTS) Deployment Handbook: Planning and Optimization Aspects. 1st ed. New York, NY, USA: Wiley, 2006.

[9] Haider B, Zafarrullah M, Islam MK. Radio frequency optimization and QoS evaluation in operational GSM network. In: WCECS 2009; 20–22 October 2009; San Francisco, CA, USA. Shanghai, China: IAENG, 2009. pp. 393–398.

[10] Francis JC, Abu El-Ata M. Benchmarking mobile network QoS. In: IEEE 2003 36th Hawaii International Conference on System Sciences; 6–9 January 2003; Waikoloa Village, HI, USA. New York, NY, USA: IEEE, 2003. pp. 1–7.

[11] Soldani D, Li M, Cuny R. QoS and QoE Management in UMTS Cellular Systems. 1st ed. New York, NY, USA: Wiley, 2006.

[12] Nipp O, Kuhn M, Wittneben A, Schweinhuber T. Speech quality evaluation and benchmarking in cellular mobile networks. In: IEEE 2007 Mobile and Wireless Communications Summit; 1–5 July 2007; Budapest, Hungary. New York, NY, USA: IEEE, 2007. pp. 1–5.

[13] Hapsari WA, Umesh A, Iwamura M, Tomala M, Gyula B, Sebire B. Minimization of drive tests solution in 3GPP. IEEE Commun Mag 2012; 50: 27–36.

[14] ETSI Technical Specification. Speech Processing, Transmission and Quality Aspects (STQ); QoS Aspects for Popular Services in GSM and 3G networks; Part 1: Identification of Quality of Service Aspects. Sophia Antipolis, France: ETSI, 2003.

[15] Zhang J, Sun J, Yang D. Application of drive test for QoS evaluation in 3G wireless networks. In: IEEE 2003 ICCT; 9–11 April 2003; Beijing, China. New York, NY, USA: IEEE, 2003. pp. 9–11.

[16] Mossavat I. A hierarchical Bayesian approach to modeling heterogeneity in speech quality assessment. IEEE T Audio Speech 2012; 20: 136–146.