Mobile Text Messaging Quality Assessment

Janusz H. Klink
Telecommunications and Teleinformatics Department
Wroclaw University of Science and Technology
Wroclaw, Poland
e-mail: janusz.klink@pwr.edu.pl

Abstract— **Telecommunication services are expected to be of good quality and offered for a reasonable price. Operators, competing strongly for customers, always present their products in the best light, and underline the highest service quality, which is often measured in incomparable circumstances, using different procedures and measurement methods. The paper presents the efforts of European standards institutions, regulators and operators in the scope of improving the provision of telecommunication services and ensuring quality. The author presents the main parameters that influence the quality of Short Message Service, which represents a wide range of text messaging services and shows the method and environment for measuring end-to-end delivery time. The measurement scenarios performed in both the real network and also a laboratory environment are presented. The results of measurements, performed in the real networks of four operators in Poland, show that the message delivery time fluctuates during the course of the day and also depends on the operator. Generally however, the short text service is of good quality and is highly assessed by users. The author also presents also the quality of experience model for text messaging.**

Keywords- mobile text messaging; SMS; QoS; QoE; quality assessment.

I. INTRODUCTION

In the last few decades, mobile telephony has reached a deeper level of penetration worldwide than cars, radio or TV. From over 700 million registered users in 2000, the mobile cellular industry has grown considerably and exceeded 7 billion subscriptions in 2015. With the increased number of mobile subscribers over the world, Short Message Service (SMS), has gained a huge popularity among the users (around 8 trillion messages a year) and is also described in numerous scientific papers [1-4]. Moreover, after voice, messaging is the biggest revenue-generating mobile service on the telecommunication market [5]. Although, in some countries SMS has peaked, and the traffic volumes are in decline, there are still more countries where overall SMS traffic and its use-per-subscriber is still growing. A significant growth in mobile subscribers is observed in the Middle East, Asia, Africa and Latin America, thus the dominance of SMS in the immediate future is unthreatened. According to [5] SMS will be one of the major worldwide communication tools for the next decade, despite the progressive extension of user equipment utility. The increase in the processing power of mobile devices has made them significantly more multi-functional and allows Internet browsing, emailing, multimedia and instant messaging. Despite the rapid growth of the so called Over-The-Top (OTT) messaging apps and Voice over IP (VoIP) services, SMS is still generating more than half of the total mobile messaging revenue (Fig. 1) [6].

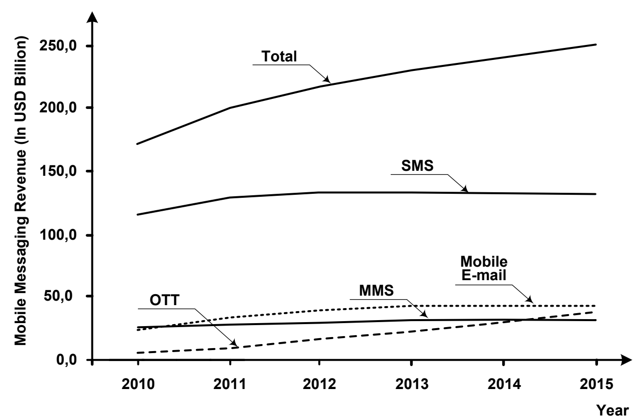

Figure 1. Mobile messaging revenue in recent years

Such large revenues mean that SMS will remain an important service for mobile operators for many years. While Instant Messaging (IM) and SMS are both text messaging services, there are differences that encourage different user behavior. IM is a two-way communication with many quick responses, whereas SMS is an individually paid-for message that is used to just send information. While there are low to zero costs for the user when using IM, travelers stick to using SMS as it is cheaper than purchasing a mobile data package or subscribing for a data roaming plan in order to send a few messages. SMS is common to all phones and almost all users [6] while IM usually requires smartphones with dedicated apps and specific knowledge of how to use it, which can be a barrier for some sections (older) of society. Moreover, the market for IM is fragmented by different services that cannot communicate with each other. Users also choose IM apps based on their geographical location. While WeChat and Line are clearly the leaders in China and Japan respectively, WhatsApp is also far bigger in East and South-East Asia than the USA.

One of the major factors that allow service providers to keep their customers is the price, which is diminishing year by year, the next factors being common availability, simplicity. and good quality. Service quality is becoming an increasingly more important factor to users at the moment of choosing a network operator or service provider [2-8]. Thus, many operators' efforts are concentrated on the efficient mechanisms for handling the message traffic [9]. On the other hand, a service-oriented management is focused on service quality rather than network performance [10]. An effective evaluation of service quality can help a service provider to increase customer satisfaction which plays a key role in influencing a user's decision on staying with or changing a provider. Soldani and al. [11] claim that over 80% of customer defections (churning) are due to frustration over the product or service and the inability of the provider/operator to deal with this effectively. Moreover, this is a chain reaction, where one frustrated customer will tell other people about his bad experience. An operator cannot assess the level of its service quality based on checking customer complaints, but should have a more active attitude as statistics have shown that for every person who calls with a problem, there are another 29 who will never call. On the other hand, about 90% of customers will not complain before defecting. This churn directly affects the profitability and image of the operator or service provider. Thus, the way to avoid such situation is to devise a strategy to manage and improve the Quality of Experience (QoE) [12].

The QoE based research mechanism for control and management of resources is also getting more attention in literature [13-15]. The SMS quality evaluation is also important in terms of the professional deployment of the service [16]. It is of great importance to the many local governments (e.g., in Poland) that implement SMS to notify inhabitants and all people about emergency states [17]. There has been a noticeable increase in text messaging applications in healthcare, as a part of mobile Health (mHealth) programs, for behavior modification, disease management and surveillance, prevention and public health education, data collection, etc. [18, 19]. On the other hand, growing competition among service providers and network operators forces these entities to provide high quality services. The question of how to describe this quality and what parameters should be used, is asked not only by operators but also by regulators of the telecommunications market in the European Union [20]. One of the factors motivating telecommunications industry operators to act in this direction are regulations undertaken both at the European level [21, 22], and in individual Member States [23, 24]. A particular example of this can be seen in the case of Poland, where on the Electronic Communications Office's initiative (2012), a Memorandum on Cooperation for Improving the Quality of Services in the Telecommunications Market has been signed. The first stage of works was finalized in the form of an official report [25], which was published by the Electronic Communications Office in February 2014. Despite long discussions, the current edition of this report does not define any QoS requirements for SMS, but there is a hope for a gradual expansion of the scope of this document. Although, SMS is not a real-time service, it is often perceived as such (a near-real-time) service [26] by a huge amount of users. Therefore, two factors seem to be important from the QoS point of view, these are: delivery rate and time of delivery. Nowadays, the delivery rate is mainly at a high level, reaching, in the case of many operators, values around 95% [17]. However, these factors are correlated with each other, because the delivery rate also depends, among other things, on the delivery threshold time after which the message is considered as lost. Therefore, delivery time seems to be a key performance indicator (KPI), which is much more crucial for the service quality perceived by the users. They want the information to be delivered in acceptable time. But what does it mean? In the era of information and communication technologies with more and more bandwidth and rich service offerings, user demands concerning the service are also growing. Today SMS is more often treated as an almost instantaneous communication medium for rapid exchange of information, and even a form of text dialogue between people [26]. The arrival delays could be a serious problem for time sensitive content such as customer account changes, last-minute tickets, product availability notifications etc. Online booking services and airlines, for example, already use SMS to notify travelers of the status of flights [27, 28]. A relatively short time of message delivery is one of the main factors describing SMS quality affecting its application and popularity among users. In Section II the author presents the basics of SMS functionality and the main parameters and statistics describing the quality of the service, according to the ETSI standards [29]. Section III describes the main prerequisites concerning service quality assessment in the real network. Sections IV and V present the method and tools used during measurements in the real network, taking into account connections inside one network and between different networks, respectively. The message delivery time distributions are also presented and discussed. In Section VI the author proposes, on the basis of measurements results, the Quality of Experience model for SMS. Section VII presents the conclusions and the plans for the future work.

II. Short Message Service Quality

SMS, belonging to the so-called "non-real-time" class, is a "store-and-forward" type of service [30]. Communication between two users is done via at least one server, acting as an intermediary unit. A user's equipment transmits a message to the server, which optionally sends it to the next server and so forth. The end server, after receiving the message, informs the recipient's equipment of receiving a message and, finally, the user can read the message. SMS was originally designed for the transmission of text information, where the length of a single data unit cannot exceed 140 bytes and, according to ETSI standards [31], remains constant regardless of the number of characters transmitted in a single message. Depending on the alphabet used, the maximum message length may vary between 70 and 160 characters. When the information is longer, then it is divided and encapsulated into several 140-byte data units, and sent as separate messages.

According to ITU-T recommendations the quality study presented in this paper only takes into account the information that does not exceed the size of a single data unit [32]. The test scenarios for concatenated message sending are for further study. On the other hand, ETSI standards give very detailed information regarding SMS quality parameters and their computation.

The most important parameters are:

- SMS Successful Ratio - the ratio of correctly sent messages, expressed by the probability of correct message sending and its delivery to a service center,
- Completion Rate for SMS - indicator of properly delivered messages, expressed by the percentage of messages successfully sent and delivered to a recipient,
- End-to-End Delivery Time for SMS - time to deliver a message from end to end, expressed as the time measured from the moment of sending a message by a sender to a service center until it is received by a recipient.

Monitoring of the parameters, mentioned above, is crucial for the operator, who has to watch over the process of service delivery at every stage of its implementation. It gives the knowledge of network performance, which in turn impacts the quality of service [33]. The measurements can be done based on the following scenarios:

- using all real traffic in the network (in a specific period of time),
- using a sample of real traffic,
- making the test calls,
- a combination of the above.

Message loss and message integrity are valid concerns, however, they are handled by lower layer network mechanisms and protocol, which are outside the scope of this paper. From the user's point of view, it is very important that messages are delivered to the recipient as soon as possible and in an unchanged form. From this perspective, it can be seen that the parameter which probably has the strongest impact on the SMS quality, perceived by the user, is the End-to-End Delivery Time.

According to the ETSI standard [29], the following statistics should be provided separately:

- the mean value in seconds for sending and receiving short messages,
- the time in seconds within which the fastest 95% of short messages are sent and received,
- the number of observations performed.

It should be noted that, concerning the mobile environment, the values of QoS parameters, can be affected not only by congestion in the SMS system or signalling channels but also by network or service non-accessibility in the claimed area of coverage. In that case, operators may wish to distinguish the effects of coverage and access congestion. From the user's point of view there is no need to do it, because all these phenomena have an impact on the end-user perception.

III. Measurements Set-Up

The most representative for the network are statistics calculated from measurements on real traffic performed in a long period of time and in the whole network. This is obviously the most expensive case and it is therefore valuable to limit the measurements to the needed number and representative population of Network Termination Points (NTPs) or Service Access Points (SAPs). When sampling or test calls are used, it should be ensured that the results reflect the service quality perceived by customers. In general, the choice of adequate origin and destination NTPs for the measurements may be based on the national/international numbering plan, on traffic patterns/distribution or on geographic coverage. Measurements should be scheduled so as to reflect traffic variations over the hours of a day, the days of the week and even the months of the year [29]. Network specific characteristics and user behaviour, depending on the kind of networks under study, i.e., fixed, mobile or a combination of them, need to be taken into account. SMS quality assessment using an intrusive method (based on test messages sending) has the advantage to precisely analyse the service quality over defined configurations because it is an end to end analysis and correlates well with user perception. The main drawback of such a kind of method is that it provides a limited view of the quality of service provided by the operator or service provider, in general. This is due to a restriction of the measurements to the analysed test configuration or specific network area. It is obviously impossible to do such kinds of measurements everywhere in the network and it is therefore important to specify the number of measurement locations, so that the measurement points can be spread according to the size of the examined network. In practice, the drive tests are used to determine the quality of service that is experienced by users in main meeting places and communication paths. As a rule of thumb, the number of samples (tests) within a measurement campaign correlates with the reliability of results. In general, the higher the number of collected data samples, the more precise and trustworthy the results are. The number of observations for message delivery times depends, like the number of observations for any quantitative variables, on the variability of measured data. Therefore there is a need to perform some pilot tests or take into consideration the results of former measurements in the network in order to get some initial data. The number of observations can be calculated by formula (1):

$$n = \frac{z^2_{1-\frac{\alpha}{2}}}{a^2}\left(\frac{s}{mean(x)}\right)^2 \qquad (1)$$

where:

$z^2_{1-\alpha/2}$ – the 1-α/2-percentile of the standard normal distribution,

s – the expected standard deviation of the call set up time (from former measurements),

mean(x) – the expected mean value of the measured value (from former measurements).

a – the relative accuracy.

The results are presented in Table I.

TABLE I. NUMBER OF OBSERVATIONS NEEDED

s/mean(x)	No. of observations
< 0.1	100
0.1 to 0.3	1 000
> 0.3 to 0.5	2 500
> 0.5 to 0.7	5 000
> 0.7 to 0.9	7 500
> 0.9	10 000

where:

$z^2_{1-\alpha/2}$ = 1.96 for a confidence level of 95%,

a = 2%,

Taking into account the number of tests to be performed and the usually limited time of measurement, particular attention should be paid to the scheduling of SMS tests. This means that a subsequent SM towards the same destination can affect the result of a previous SMS test if the second SM is sent before stating the result of the first one [31]. So, to avoid such a situation, a proper time interval between the sending of messages should be kept. An alternative way to increase the number of tests is by sending the SMs towards different destinations. In such a way, the queuing per destination mechanism is bypassed.

IV. IN-NET MEASUREMENTS PERFORMED IN THE REAL ENVIRONMENT

This Section presents the method, measurement environment, results, and evaluation of SMS quality provided by leading mobile operators functioning in Poland, i.e., Orange, Play, Plus and T-Mobile.

A. Method and the Measurement Environment

Data was collected from more than 120 000 tests (individual observations) performed during one week in Wroclaw - one of the biggest cities in Poland (650 000 inhabitants). The test environment (Fig. 2) consists of measuring robots (one per operator), each covering a Personal Computer with a 3G modem and specially designed application that manages the measurement and data collection process. Each robot plays both roles: sender and receiver. Initially, the first one sends a previously prepared text message to the Short Message Service Center (SMSC) located in the operator network, inserting its own number (i.e., both: the sender and receiver belong to the same network) in the destination address field. The time of the message sending is written down into a special record of a log file on the robot's hard disk. The SMSC then sends the message further, i.e., to the receiver, which in this situation is the same robot that sent the message before. Next, the receiving part of the robot is informed of the incoming message and then it also records the current time in the log file. In this way, the file collects a number of records with the times of sending and receiving the particular messages. In the next step the robot sends the message to another robot that is connected to the other operator's network. All previous actions are repeated and then the next operator's network is chosen as the destination in a round-robin scheme. The robot software allows the setting of the frequency of message sending in an order, based on message delivery time observed, in order to avoid traffic congestion (Fig. 3).

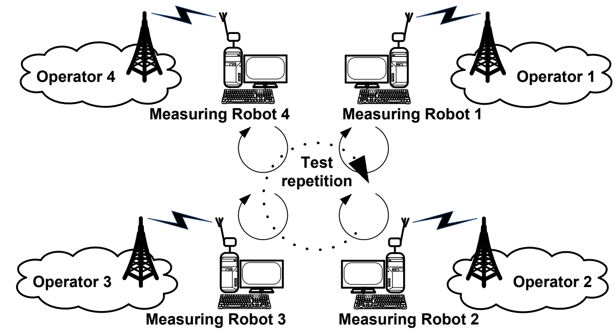

Figure 2. The test environment for the SMS parameter measurement in real mobile networks

It should be noted that such a solution causes the risk of substantially reducing the number of tests when the delivery time increases enormously. In order to eliminate such phenomenon, it is possible to limit the maximum acceptable delivery time and, if it is overrun, the expected message may be recognized as lost. Then the sending of the next message can start. In this phase of the measurements such a time limit was not used so that the robots were able to capture all delivery time cases.

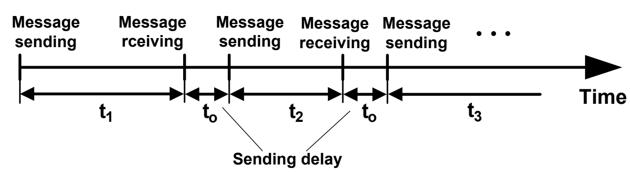

Figure 3. Sequential message sending scenario

B. Message Delivery Time Distributions

The measurements performed in real networks, as mentioned before, allowed the message delivery time distributions in each network to be determined (Figs. 4-7). After analysis of the distributions it can be stated that very long delivery times occur in the case of some operators (e.g., Play), which may be irritating to some customers. From a statistical point of view, however, they are not of considerable importance (it concerns about 10^{-4} cases). Moreover, it is negligible especially in the case of discarding 5% of the highest values before further analysis. The majority of the captured message delivery times do not exceed 10 seconds which means that SMS users should be satisfied. Moreover, almost 99% of the messages were delivered in a time of no higher than 6 seconds in the case of three operators. Only in the Play network the message delivery time distribution was different. The question is: are the captured message delivery times satisfactory in the case of real time text messaging applications? More detailed analysis will be shown in the Section IV, where the scores of subjective measurements will be taken into account.

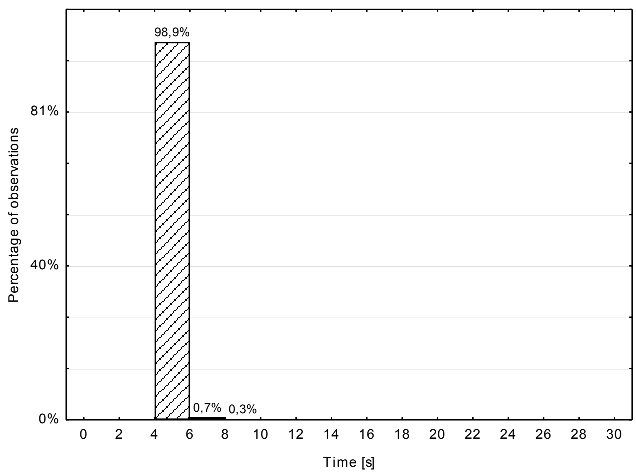

Figure 4. Message delivery time distribution in the Orange network

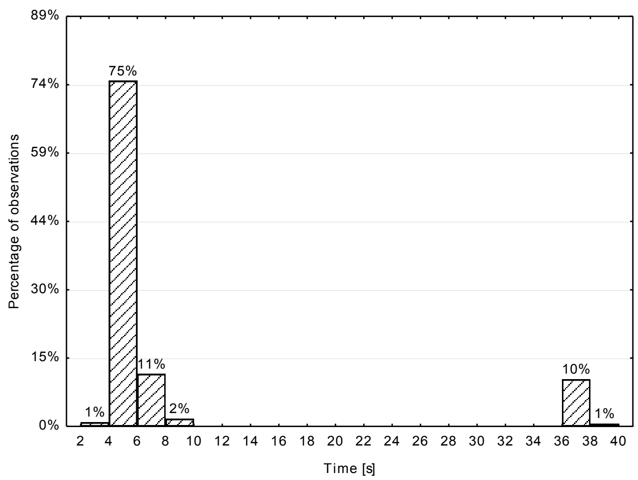

Figure 5. Message delivery time distribution in the Play network

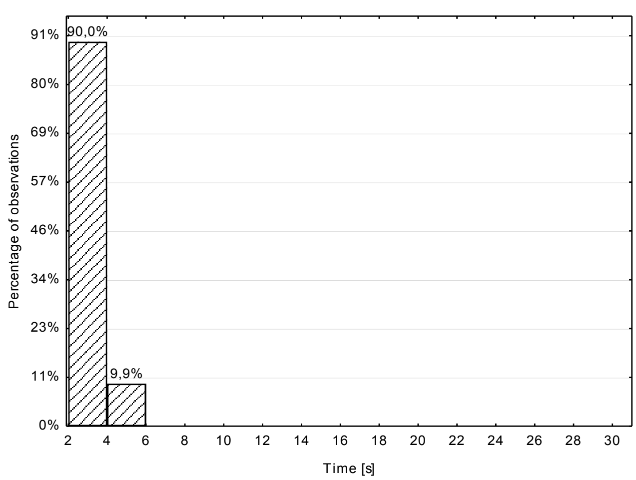

Figure 6. Message delivery time distribution in the Plus network

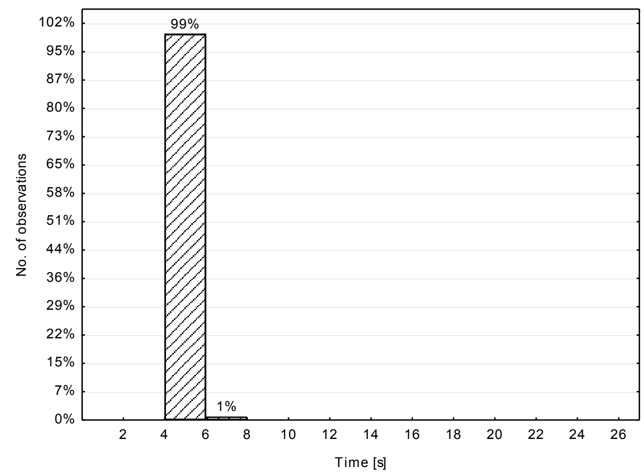

Figure 7. Message delivery time distribution in the T-Mobile network

C. *Message Delivery Time as a Function of the Time of Day*

Analysis of the results showed that message delivery time, as expected, is not stable and is diversified depending on the operator (see Table II).

It can be noted that the lowest mean value of the message delivery time (3.6 s) occurred in the case of operator No. 3.

TABLE II. COMPARISON OF MESSAGE DELIVERY TIMES

No.	Operator (number)		Delivery time [s]				
			Mean	Min.	Max.	Std. dev.	Median
1	Orange	(OR)	4.66	2.8	26.1	0.47	4.6
2	Play	(PY)	15.79	3.1	**10815**	**256.4**	4.5
3	Plus	(PS)	**3.6**	3.3	65.7	0.68	3.5
4	T-Mobile	(TM)	4.96	3.1	23.7	0.39	4.9

A slightly worse score can be seen in networks 1 and 4. The longest delivery times, as well as standard deviation, was offered by operator No. 2. Such rough analysis can lead us to the conclusion that the SMS does not work properly and many of users may be dissatisfied with the service. On the other hand, when we take into account the median, which is by definition the value located in the middle of the population, it can be noted that it is comparable with the appropriate parameters of the other operators. Moreover, the median value seems to be a better parameter describing the service quality experienced by the users in the case of high standard deviation of QoS parameters. As mentioned before, ETSI proposes to describe the message delivery time by presenting the time within which the fastest 95% of short messages are sent and received [29]. According to the above, the distributions of message delivery times as a function of the time of day are presented in Figs. 8-11. The black points represent median values, whereas the dashed boxes show the ranges of delivery time after discarding 5% of the lowest and highest values, respectively. In other words, they represent 90% of the population. Moreover, the top level of each dashed box

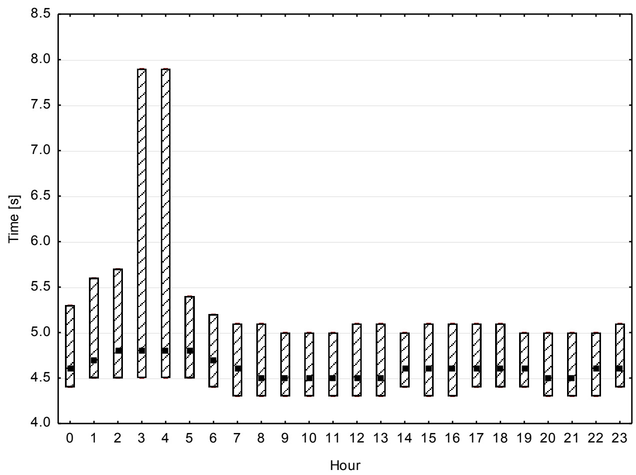

Figure 8. Short message delivery time in the Orange network as a function of the time of day

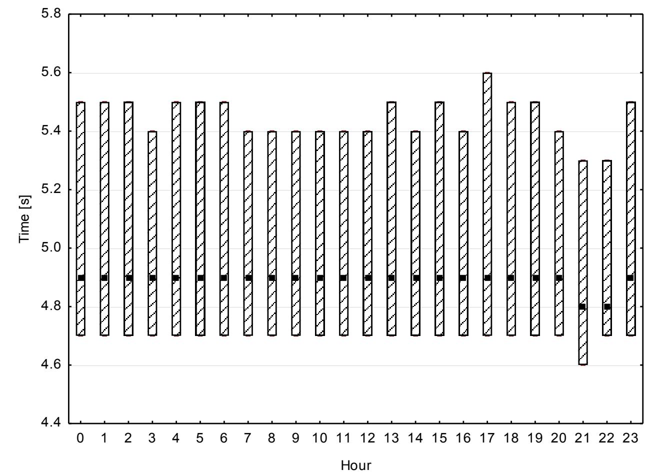

Figure 11. Short message delivery time in the T-Mobile network as a function of the time of day

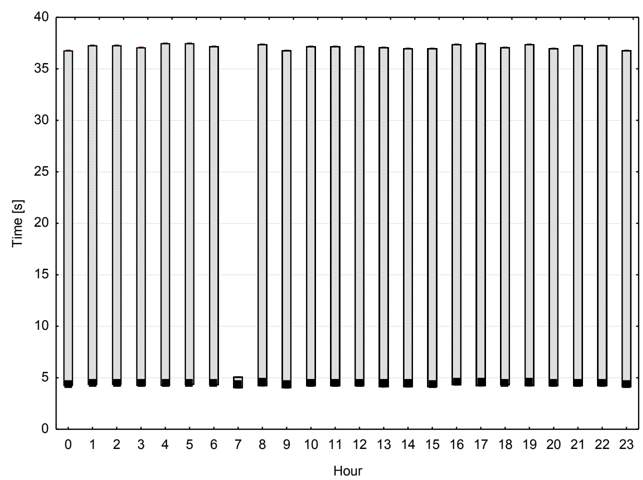

Figure 9. Short message delivery time in the Play network as a function of the time of day

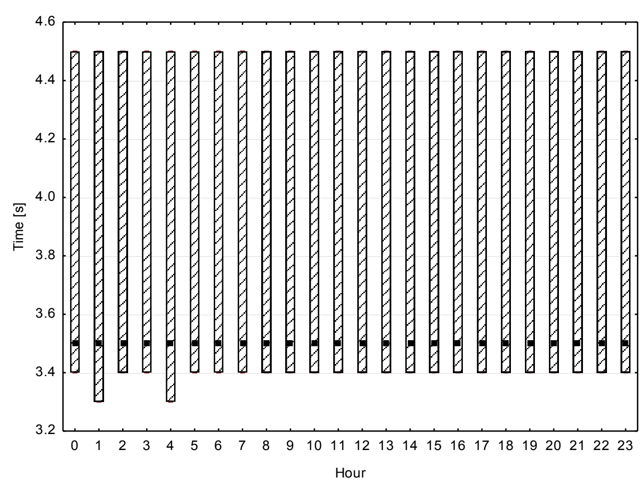

Figure 10. Short message delivery time in the Plus network as a function of the time of day

denotes the highest limit of delivery time for 95% of messages sent in the relevant hour [29]. As presented in Fig. 4, fluctuations of the message delivery time experienced by almost all SMS users of the Orange network, do not exceed the value of 6 seconds. The highest deviation is observed at around 3 and 5 o'clock in the morning. The results of the observations performed in the Play network show the values of delivery time deviation that are almost at the same level, except for one hour (7 a.m.), during the whole day. However, compared to the Orange network, the range of deviations, observed in the individual time intervals, is much wider here. Moreover, it can be seen that users of the Play network experience relatively high fluctuations of message delivery times not only in specific hours but during the whole day. It should be noted that the median values and deviations of the message delivery times, presented in Figs. 8- 11, are valid for 90% of observations and may slightly differ from the values shown in Table I, which takes into account all the captured data. Although the median (or even the mean) values of the message delivery times and their deviations can be used to compare the different operator's network performances or QoS parameters, they do not answer the question concerning the quality assessed by the users. For this reason, the relation between objectively measured QoS parameters and the quality of experience (QoE), which is subjective, should be determined.

V. Inter-Net Measurements performed in the Real Environment

In this section, the author presents the second part of the measurements performed in the real network. This is an extension of the previous investigations and not only covers separate measurements performed in networks of different operators (In-Net) but also takes into account relations between networks (Inter-Net) as well. Although the operators remained the same, the author changed the location of the measurements from an urban to rural area of the same part of Poland (Lower Silesia district).

A. Method and the Measurement Environment

Each "inter-operator" relation is described by 1440 measurements, i.e., one measurement per minute (60 samples per hour), which results in more than 23 000 measurements in total. The test environment consists of measuring robots (one for the operator), analogous to the ones presented in Fig. 2. However, the measurement scenario was extended to inter-network investigations. In order to compare Inter-Net relations, the author, as a first step, performed the In-Net investigations for each network separately. The author does not analyse delivery times as a function of the time of day but shows how the examined networks fulfil ETSI requirements [29, 30]. One of the main parameters, as mentioned before, is the time in seconds within which the fastest 95% of short messages are sent and received. Therefore, the cumulative distribution function (CDF) of the message end-to-end delivery times is determined for each operator.

Figure 12 shows that at least 95% of messages in the examined (rural) area of the Orange network are delivered in a time of no longer than 16 s. That is more than twice as long as in the urban area tested before (see Fig. 4). Moreover, as presented in Fig. 13, these times are of higher deviation and most of the messages are delivered between the 10th and 12th second. Significantly longer delivery times were observed in the Play network, where 95% of messages were delivered to their destination in a time exceeding 30 seconds (Fig. 14).

It is interesting that most of the messages reach the receiving station, as in the Orange network, in comparable time, i.e., in less than 10 seconds. Twelve seconds are needed to deliver 84% of the messages. However, very long delivery times are observed here more often than in the case of measurements in the urban area. As presented in Fig. 15, over 7% of messages are delivered in a time exceeding 30 seconds.

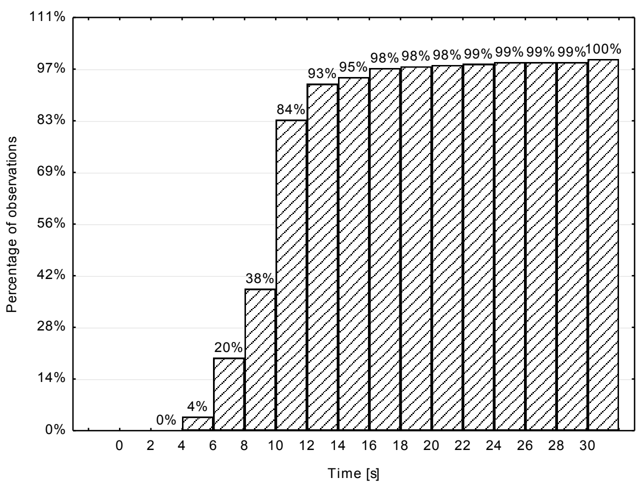

Figure 12. Cumulative Distribution Function of end-to-end message delivery time in the Orange network

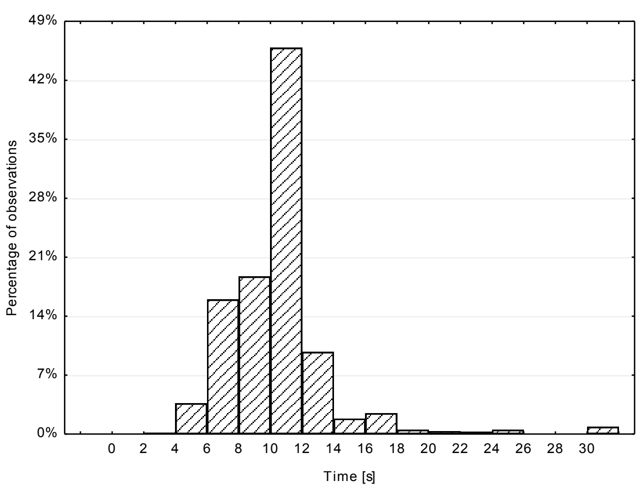

Figure 13. Probability Density Function of end-to-end message delivery time in the Orange network

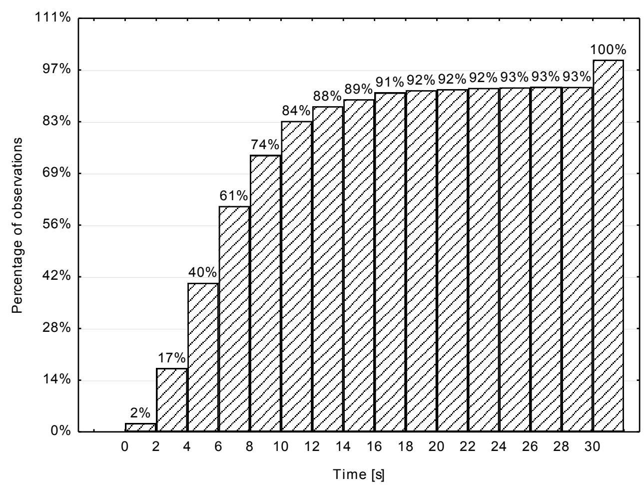

Figure 14. Cumulative Distribution Function of end-to-end message delivery time in the Play network

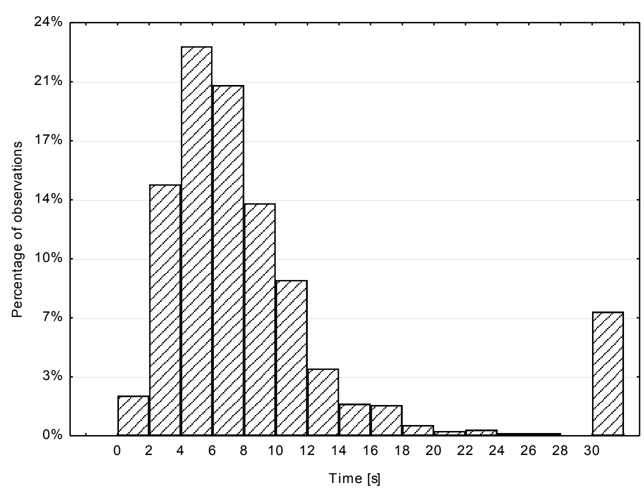

Figure 15. Probability Density Function of end-to-end message delivery time in the Play network

The fact that the vast majority of messages in the Play network are delivered in a relatively short time can positively influence user perception of the service. On the other hand, a fairly large amount of messages are delivered in a time of longer than 30 s, which means that 95% of them are delivered in the longest time when compared with the other examined networks.

The Plus and T-Mobile networks behave in similar way to the Orange network, which means that 95% of messages are delivered in a time of less than 20 seconds (in the case of the last two it is 18 s) and the observed delivery times are of a standard deviation that is significantly lower than in the Play network (Fig. 16-Fig. 19).

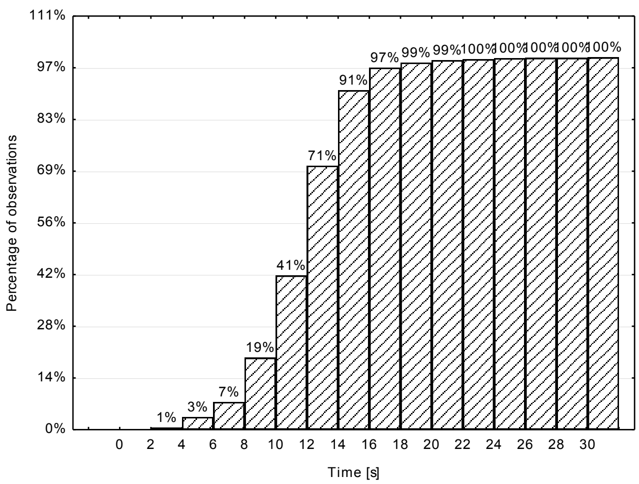

Figure 16. Cumulative Distribution Function of end-to-end message delivery time in the Plus network

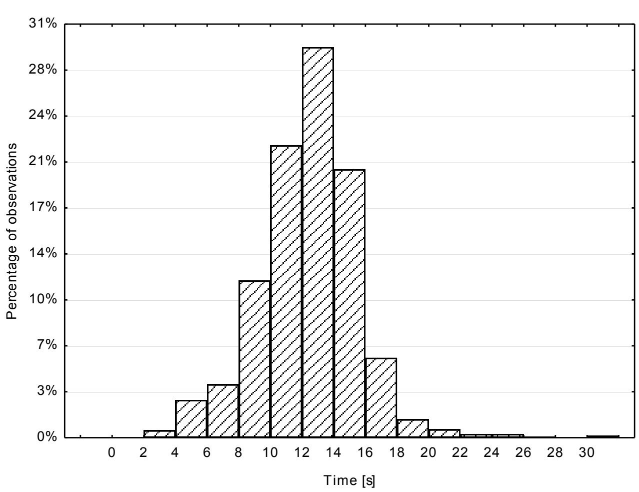

Figure 17. Probability Density Function of end-to-end message delivery time in the Plus network

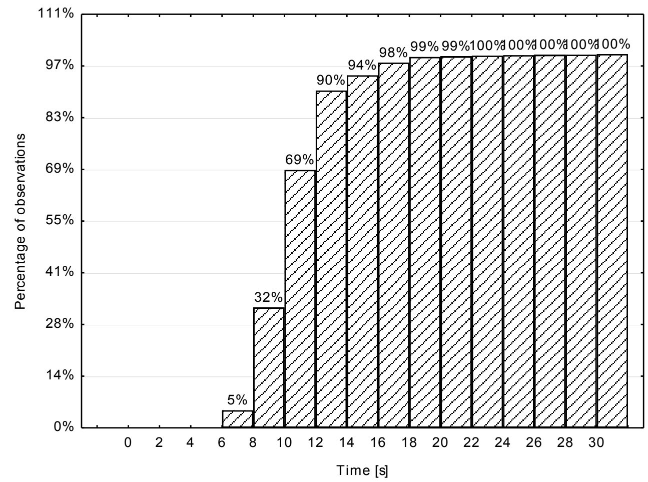

Figure 18. Cumulative Distribution Function of end-to-end message delivery time in the T-Mobile network

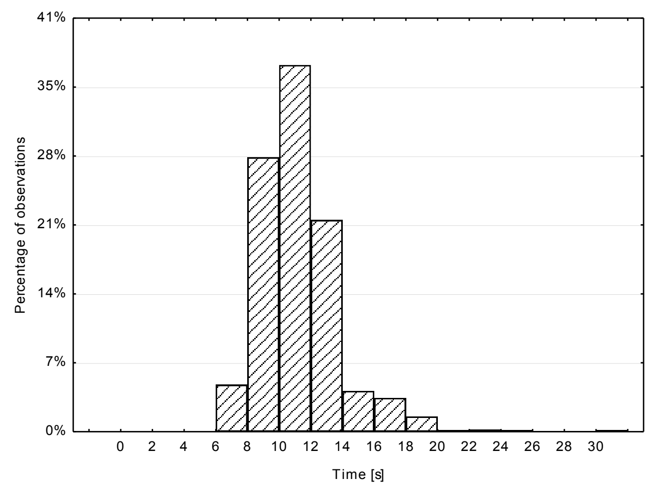

Figure 19. Probability Density Function of end-to-end message delivery time in the T-Mobile network

Analysis of the message delivery process shows that the mean value of message delivery times, in the case of messages exchanged between users belonging to different operator networks, is usually higher than in the case of transmitting them inside an individual network. Exceptions to this rule are seen in the Play and Orange networks where In-Net measurements show comparable or even higher mean values of message delivery times (Tables III and IV). In almost every network the maximum values of message delivery times significantly exceed the mean values (even 20 times or more). Additionally, high values of standard deviation of delivery times are observed for almost all the operators. An exception to the rule is the Plus network, where these times are low and rather stable, i.e., of low standard deviation. It concerns both, In-Net and Inter-Net measurements as well.

TABLE III. COMPARISON OF MESSAGE DELIVERY TIMES INSIDE NETWORKS

Operator		Delivery time [s]					
		Mean	Min.	Max.	Std. dev.	Median	Skew
Orange	(OR)	11.5	3	312	10.2	11	20.5
Play	(PY)	41.1	1	**2170**	**201.3**	7	7.7
Plus	(PS)	13	3	74	3.8	13	5.6
T-Mobile	(TM)	11.8	7	59	2.9	11	4.7

TABLE IV. COMPARISON OF MESSAGE DELIVERY TIMES BETWEEN NETWORKS

Operator		Delivery time [s]					
		Mean	Min.	Max.	Std. dev.	Median	Skew
OR	PY	18.4	3	563	42.6	11	8.1
OR	PS	19.6	1	1046	75.4	9	9.9
OR	TM	11.2	4	244	8.9	10	17.1
PY	OR	107.3	1	3268	414.5	8	5
PY	PS	37.5	1	2154	200.4	6	7.7
PY	TM	41.5	1	2186	209.8	7	7.2
PS	OR	14	5	128	5	14	11.4
PS	PY	15.4	6	196	9.7	15	10.5
PS	TM	13.9	5	74	3.4	14	3.5
TM	OR	13	7	46	2.7	13	2
TM	PY	127	7	4293	539.4	13	5.7
TM	PS	13.3	6	202	11.1	11	8.1

For all the In-Net and Inter-Net relations a median value of message delivery time was determined. As presented in Tables III and IV, in most cases and due to asymmetry of its probability distributions, these values significantly differ from the mean values. Analysis of the data shows non-zero (positive) skewness, which denotes that most of the delivery times measured are lower than the calculated mean value. It suggests that the median value should not be treated as the main performance indicator, describing quality of the service.

In general, the median is better suited for skewed distributions, like in our cases, to derive a central tendency since it is much more robust and sensible. If the median is not equal to the mean, it is an indicator for statistical outliers, i.e., a small number of extreme measurement data values that significantly influence the mean value [34]. From a statistical point of view it is understandable, but from a practical point of view the median value additionally gives us information about the so called common user perception of the service quality. Even if the mean value of message delivery time is high, a median value could be quite moderate, which denotes that the vast majority of the users are satisfied with the service quality.

As mentioned before, from a practical point of view, the message delivery time is the most relevant statistic for at least 95% of messages [29]. The results are presented in Table V.

More detailed analysis of the longest delivery times for 95% of the fastest messages, with the participation of the stations connected to the Play network, showed that these times are as follows:

- 75 s – for messages inside the Play network,
- 560 s – for messages sent from the Play to Orange network,
- 240 s – for messages sent from the T-Mobile to Play network.

TABLE V. COMPARISON OF MESSAGE DELIVERY TIMES FOR 95% OF THE FASTEST MESSAGES

Originating network	Destination network / Max. delivery times for the fastest 95% of messages [s]			
OR	**OR**	**PY**	**PS**	**TM**
	16	28	18	16
PY	**OR**	**PY**	**PS**	**TM**
	>30 (560)	**>30 (75)**	18	16
PS	**OR**	**PY**	**PS**	**TM**
	20	20	18	18
TM	**OR**	**PY**	**PS**	**TM**
	18	**>30 (240)**	20	18

Such long delivery times are usually acceptable for general-purpose use, but may be critical or even unacceptable for professional applications and near-real-time communication.

VI. QOE MODEL

This paragraph presents the method and test-bed for assessing the quality experienced by users (QoE) of text messaging services. When identifying QoE metrics in general, there are as many different expectations as there are users, but most of these expectations can be grouped under two main categories: *reliability* and *quality* [11]. The main KPIs, in the service reliability dimension, are:

- availability (anywhere),
- accessibility (anytime),
- access time (setup time),
- continuity (service retainability).

The quality category can be described by the second group of KPIs as follows:

- quality of session,
- bitrate,
- active session throughput,
- system responsiveness,
- packet loss,
- end-to-end delay and delay variation.

Depending on the type of service, the value of each of these metrics translates to a different level of impact on the actual QoE. Some can be totally irrelevant in one case while being the most important in another. The most important KPIs for SMS users seem to be end-to-end delay and its variation.

The concept of the measurement environment is based on the server emulating the service provider and also several test positions representing user terminal equipments (Fig. 20). Each position consists of a Personal Computer (PC) with a special application emulating the mobile phone. A group of testers (40 selected users) send and receive short messages (1200 in total). In the experiment all test messages have the same format and content.

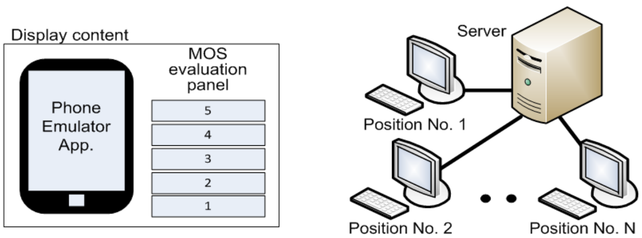

Figure 20. The laboratory test environment for the evaluation of text messaging Quality of Experience

The users send a number of messages which are passed through the server to the destination addresses after a period of time. This may be controlled during the test. After receiving the message, users assess the service quality by evaluating the end-to-end delivery time and choosing the appropriate marks from the MOS evaluation panel, where MOS stands for Mean Opinion Score, expressed in a 5-level scale (0 – the worst case and 5 – the best one, respectively). All the test parameters and user marks are stored on the server and saved for further analysis. Several dozen users took part in the experiment and more than 1.2 thousand tests were performed. The results of the tests allowed a QoE model to be built (2), which indicates a relationship between end-to-end message delivery time and service quality perceived by the users. Statistical analysis shows a significant correlation (almost 80%) between message delivery times and the users' evaluation grades. Next, regression analysis was performed and, using ordinary least squares (OLS) estimation, the approximate relation between message delivery time and users' grades (in MOS scale) was determined:

$$\mathrm{MOS} = -0.1 \cdot T + 4.97 \qquad (2)$$

where: T - message delivery time.

Due to the fact the distribution of the data was not normal (checked by the Shapiro-Wilk test [35]), the author made a validation of the model, using the Mann–Whitney–Wilcoxon (MWW) test [36]. The test can even be applied on unknown distributions, contrary to the t-test, which can only be applied on normal distributions [37]. This test showed a good estimation of the users' quality perception, under assumption of 95% confidence interval (significance level $p<0.05$).

As was mentioned before, the mean value of the QoS parameter may not always be the best indicator of the network performance or the quality perceived by users. Therefore, the author presented the SMS QoE model, which shows the relation between the message delivery times and the median values of user ratings (Fig. 21). Thus, the author proposed a new name for the scale, i.e., Median Opinion Score (MedOS). The black points represent the median value, whereas the upper level of the dashed boxes determines the scores given by the 95% of users experiencing the specific message delivery times. Four levels of quality, acceptable by users, were defined, i.e., excellent quality (EQ), good quality (GQ), fair quality (FQ) and poor quality (PQ), respectively.

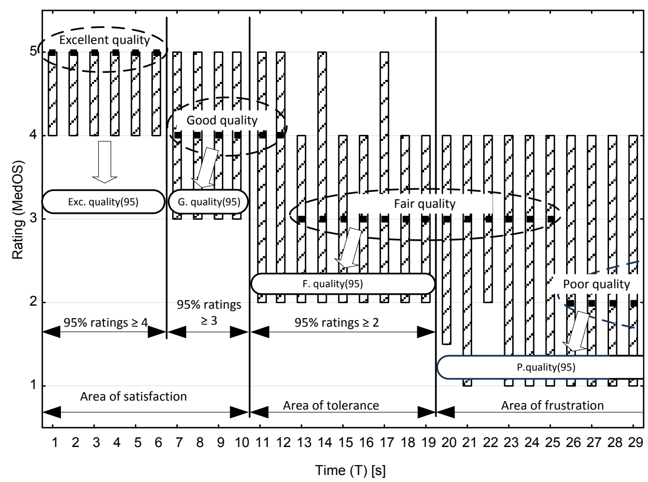

Figure 21. The quality experienced by the users majority (MedOS scale)

Next, four ranges of the message delivery times were assigned to the proper quality levels (EQ, GQ, FQ or PQ) on the basis of median values of user scores, given for those times (see the dashed ellipses in Fig. 21). According to ETSI [29], the quality levels should take into account the best 95% of samples (here messages). The Excellent Quality level, denoted by EQ(95), is reached when the median of user scores is equal to 5, and 95% of the samples have ratings equal or higher than 4.

TABLE VI. SHORT MESSAGE DELIVERY TIME DISTRIBUTION FOR 95% OF CAPTURED SAMPLES – IN-NET URBAN AREA

Operator	Percentage (*P*) [%]			WAQF
	EQ	GQ	FQ	
OR	100	0	0	5
PY	80	13.7	6.3	4.74
PS	100	0	0	5
TM	100	0	0	5

The same procedure is applied to the other quality levels (there were no scores falling into the poor quality range). In this way, the relations between the message delivery times (QoS) and proper quality levels (QoE) were determined. Table VI presents, according to the MedOS scale, the measurement results obtained from the four examined real networks. Next, the Weighted Average Quality Factor (WAQF) was calculated using (3).

$$\text{WAQF} = (P_{\text{EQ}} \cdot 5 + P_{\text{GQ}} \cdot 4 + P_{\text{FQ}} \cdot 3 + P_{\text{PQ}} \cdot 2)/100 \quad (3)$$

The measurements in the rural area revealed that quite a significant percentage of messages that were delivered in very long times, recognized as unacceptable in near-real-time communication (see Tables VII and VIII).

TABLE VII. SHORT MESSAGE DELIVERY TIME DISTRIBUTION FOR 95% OF CAPTURED SAMPLES – IN-NET, RURAL AREA

Operator	Percentage (*P*) [%]				WAQF
	EQ	GQ	FQ	PQ	
OR	4	34	60	2	2.23
PY	40	34	18	8	3.73
PS	3	16	80	1	1.64
TM	0	32	67	1	2

TABLE VIII. SHORT MESSAGE DELIVERY TIME DISTRIBUTION FOR 95% OF CAPTURED SAMPLES – INTER_NET, RURAL AREA

Relation	Percentage (*P*) [%]				WAQF
	EQ	GQ	FQ	PQ	
OR-PY	1	48	45	7	2,59
OR-PS	2	68	26	4	3,19
OR-TM	2	52	45	1	2,68
PY-OR	39	30	19	12	3,61
PY-PS	54	30	11	5	4,14
PY-TM	41	38	17	4	3,85
PS-OR	1	11	87	1	1,41
PS-PY	1	9	88	2	1,36
PS-TM	2	11	86	1	1,45
TM-OR	0	14	85	1	1,46
TM-PY	0	21	60	19	1,85
TM_PS	1	39	55	5	2,29

In this case the WAQF differs considerably from the In-Net results obtained in the urban area, which were higher. On the other hand, there was no significant difference between In-Net and Inter-Net results observed in the rural area.

WAQF seems to be used as one of the parameters that allow a comparison of the SMS quality provisioned by different operators. However, it is necessary to be very careful when drawing general conclusions concerning QoE in the examined networks based on this factor due to the fact there is no simple relation between these measures. This should be a matter of further investigation.

More detailed statistical analysis seems to be good in such a case and the median value and the time within which the fastest 95% of short messages are sent and received [29] should at least be taken into account.

VII. CONCLUSION AND FUTURE WORK

Nowadays, the text messaging is one of the most popular means of communication and a high quality of the service is therefore crucial in today's competitive market. Operators should continuously monitor network performance parameters in order to detect and isolate the problems and different kinds of threats that can impact on the quality experienced by the end-users. Thus, it is very important to not only have knowledge about the values of objectively measured performance parameters, but also about their influence on the service quality subjectively perceived by users. The results presented in the article show that the SMS provisioned by the operators functioning in the examined area of the Polish telecommunication market is of very good quality and can be used, to some extent, as a medium which also supports other kinds of text communication, especially those that require short end-to-end delivery times and immediate user-to-user interactions. Obviously, message delivery time fluctuates during the course of the day and also depends on the operator, but generally brings great satisfaction to users. It should also be noted that such a relatively small amount of collected data does not allow to a general statement about the whole Polish network to be made. Such a generalization could be made after collecting data from a bigger and representative number of selected areas, which will be done in the next step of the investigations. The author plans to build a Service Quality Monitoring System, which will be prepared to measure SMS KPIs in a wider area using so called drive tests and a central management system that will control the measurement scenarios and data collection. All data will be centrally processed and presented on a web-page.

It should be underlined that although SMS cannot be treated as a real-time messaging service, it can in some cases be used as an alternative. The main strengths of SMS are worldwide availability and the fact that there are no special requirements for user equipment or any specific software applications. Further work will be devoted to developing the QoE model towards a more comprehensive investigation of the quality issues, regarding not only intra- but also inter-operator communication.

ACKNOWLEDGMENT

The author would like to thank Pawel Bardowski for his contribution to the test-bed set-up and his valuable remarks in the course of preparing this article. A special thanks to Zbigniew Salamacha for his work during the editorial process.

REFERENCES

[1] J. Klink and P. Bardowski, "The methodology of text messaging quality assessment," ComputationWorld 2016. The Eighth International Conferences on Advanced Service Computing. Service Computation 2016 IARIA, March 20 - 24, 2016.

[2] O. Turel, A. Serenko, and N. Bontis, „User acceptance of wireless short messaging services: Deconstructing perceived value," Information & Management 44, pp. 63–73, 2007, DOI: 10.1016/j.im.2006.10.005.

[3] Portio Research. *SMS: the language of 6 bilion people*. [Online]. Available from: http://www.openmarket.com/wp-content/uploads/2015/10/ OpenMarket-Portio-Research-SMS-the-language-of-6-billion-people.pdf.

[4] A. Perez, Short Message Service, in VoLTE and ViLTE: Voice and Conversational Video Services over the 4G Mobile Network, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016. doi: 10.1002/9781119318453.ch10.

[5] PortioResearch. *Worldwide A2P SMS Markets 2014-2017. Understanding and analysis of application-to-person text messaging markets worldwide.* [Online]. Available from: http://www.strikeiron.com/wp-content/uploads/2014/04/ worldwide-sms-markets-portio-strikeiron.pdf.

[6] Portio Research. Mobile Messaging Futures 2014-2018. [Online]. Available from: http://www.portioresearch.com.

[7] D. Seo, C. Ranganathan, and Y. Babad, "Two-level model of customer retention in the US mobile telecommunications

service market," Telecommunications Policy 32, 2008, pp. 182-196.

[8] R. Virvilaite, V. Saladiene, and D. Skindaras, "The Relationship between Price and Loyalty in Services Industry," Engineering Economics, Vol. 63, No. 4 (2009), pp. 96-104.

[9] J. Zheng and E. Regentova, „QoS-based Optimal Buffering for Short Message Transfer in GPRS/UMTS Networks," Vehicular Technology Conference, 2005, 2005 IEEE 61st (Volume:5), pp. 2721-2725, doi: 10.1109/VETECS. 2005.1543841.

[10] Y. Du, W. Zhou, B. Chen, and J. Song, "A QoE based Evaluation of service quality in wireless communication Network," International Conference on New Trends in Information and Service Science, 2009, pp. 552-557, doi: 10.1109/NISS.2009.163.

[11] D. Soldani, M. Li, and R. Cuny, "QoS and QoE Management in UMTS Cellular Systems," Wiley, New York 2006.

[12] H. A. Tran and A. Mellouk, "QoE Model Driven for Network Services," Wired/Wireless Internet Communication, 8th International Conference, Lulea, Sweden, June 2010.

[13] H. Yamada, N. Fukumoto, M. Isomura, S. Uemura, and M. Hayashi, "A QoE based service control scheme for RACF in IP-based FMC networks," E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services, 23-26 July 2007, Page(s):611 – 618.

[14] B. De Vleeschauwer, F. De Turck, B. Dhoedt, P. Demeester, M. Wijnants, and W. Lamotte, "End-to-end QoE Optimization Through Overlay Network Deployment," International Conference on Information Networking, 2008, 23-25 Jan. 2008 Page(s):1 – 5.

[15] J. M. Monteiro and M. S. Nunes, "A Subjective Quality Estimation Tool for the Evaluation of Video Communication Systems," Computers and Communications, 2007. ISCC 2007. 12th IEEE Symposium on 1-4 July 2007 Page(s): MW - 75 - MW – 80, 2007.

[16] A. Waadt, G. Bruck, P. Jung, M. Kowalzik, and T. Trapp, "QoS Monitoring for Professional Short-Message-Services in Mobile Networks," 2nd International Symposium on Wireless Communication Systems, Siena 2005. doi: 10.1109/ISWCS.2005.1547692.

[17] X. Meng, P. Zerfos, V. Samanta, S. H. Y. Wong, and S. Lu, "Analysis of the Reliability of a Nationwide Short Message Service," 26th IEEE International Conference on Computer Communications, INFOCOM 2007, doi: 10.1109/INFCOM.2007.211.

[18] L. Schilling, G. Bennett, S. Bull, A. Kempe, M. P. Wretling, and E. Staton, "Text Messaging in Healthcare Research Toolkit," School of Medicine, Center for Research in Implementation Science and Prevention, Unoiv. Of Colorado, October 2013. [Online]. Available from: http://archive.careinnovations.org/programs-grants/texting/.

[19] P. Mechael, H. Batavia, N. Kaonga, S. Searle, A. Kwan, A. Goldberger, L. Fu, and J. Ossman, "Barriers and gaps affecting mHealth in low and middle countries: policy white paper," Colombia, USA. Centre for Global Health and Economic Development Earth Institute, Colombia University, 2010. [Online]. Available from: http://www.mhealthknowledge.org/resources/barriers-and-gaps-affecting-mhealth-low-and-middle-income-countries.

[20] BEREC Guidelines for QoS in the scope of Net Neutrality, Ref.: X/2012/060 – 27/07/2012. [Online]. Available from: http://www.beuc.eu/ publications/2012-00510-01-e.pdf.

[21] Directive 2009/140/EC of the European Parliament and of the Council. Official Journal EU L.337/37, 25 November 2009.

[22] Directive 2009/136/EC of the European Parliament and of the Council. Official Journal EU L.337/11, 25 November 2009.

[23] The Telecommunication Law, Journal of Laws of the Republic of Poland No. 171, item 1800 with subsequent ammendments, 16 July 2004,.

[24] Regulatory Strategy of the President of the Office of Electronic Communications until 2015. [Online]. Available from: https://www.uke.gov.pl/strategia-regulacyjna-do-roku-2015-8802.

[25] Memorandum on cooperation with the aim of telecommunication service quality improvement report, February 2014. [Online]. Available from: http://www.uke.gov.pl/files/? id_plik=14759.

[26] T. L. Lai, „Service Quality and Perceived Value's Impact on Satisfaction, Intention and Usage of Short Message Service (SMS)," Information Systems Frontiers 2004, Volume 6, Issue 4, pp 353-368.

[27] A. Dickinger, P. Haghirian, J. Murphy, and A. Scharl, "An Investigation and Conceptual Model of SMS Marketing," Proceedings of the 37th Hawaii International Conference on System Sciences, 2004.

[28] F. Newell and K. N. Lemon, "Wireless Rules - New Marketing Strategies for Customer Relationship Management," New York: McGraw-Hill, 2001.

[29] ETSI EG 202 057-2 V1.3.2, User related QoS parameter definitions and measurements. Part 2: Voice telephony, Group 3 fax, modem data services and SMS, 2011.

[30] ETSI TS 102 250-2 V2.2.1, QoS aspects for popular services in mobile networks; Part 2: Definition of Quality of Service parameters and their computation, 2011.

[31] ETSI TR 102 529 V1.1.1, SMS Testing Guidelines. Measurement Methodologies and Quality Aspects, 2009.

[32] ITU-T E.804, Quality of telecommunication services: concepts, models, objectives and dependability planning – Terms and definitions related to the quality of telecommunication services. QoS aspects for popular services in mobile networks, February 2014.

[33] ITU-T E.800, Quality of telecommunication services: concepts, models, objectives and dependability planning (…). Definitions of terms related to quality of service, 2008.

[34] B. Staehle, A. Bizenhoefer, D. Schlosser, and B. Boder, "Quantifying the Influence of Network Conditions on the Service Quality Experienced by a Thin Client User," Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB), 2008 14th GI/ITG Conference, pp. 1-15.

[35] N. Razali and Y. B. Wah, "Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests," Journal of Statistical Modeling and Analytics 2 (1): 21–33, 2012.

[36] M. P. Fay and M. A. Proschan, "Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules," Statistics Surveys 4: pp. 1–39, 2010. DOI:10.1214/09-SS051.

[37] J. A. Rice, "Mathematical Statistics and Data Analysis," Third Edition, Duxbury Advanced, 2006.