Influence of Atmospheric Parameters on Speech Quality in GSM/UMTS

Miroslav Voznak and Jan Rozhon

***Abstract*—**The paper deals with a relation between atmospheric conditions and speech quality in GSM/UMTS networks. The results are based on more than 20 thousand measurements. Authors developed monitoring tool which carried out a call every five minutes and the transmitted calibrated speech samples were compared with the received by PESQ method, the computed MOS to every call was stored in database. The meteorological station, which was located in university campus provided information about a temperature, a humidity, a dew point, a rain, a wind speed and an atmospheric pressure. The aim of our research project was to investigate a correlation between the speech quality in GSM/UMTS and meteorological data. The measured data were analysed by K-means clustering method. We observed nearly a 50 percent decrease of MOS during a heavy rain. The paper describes the way of measurements and analysis of collected data.

***Keywords*—**GSM, UMTS, MOS, PESQ, Speech quality, P.862.

I. Introduction

THE speech quality quality becomes an issue these days mainly because of the transition towards next generation networks [1], [2]. GSM networks suffer from speech quality loss, although the sophisticated algorithms are used to prevent the cumulative loss of information there are still inherent disadvantages of the networks that influence the speech quality and cause the difference in speech quality in every single call. Among these disadvantageous factors we can count the location of the user with his cell phone (open space, building…), the movement speed of the user causing the Doppler Effect to take place, or distortion from the switching between base stations to appear. In addition we need to include the effect of distortion caused by the devices working on similar frequencies to those used in GSM or signal scattering.

Through years of successful expansion GSM and UMTS technologies have become a commonplace and paved their way to every aspect of human activity. We can now see the cell phones being used by almost all people to communicate, exchange messages or even access the Internet. Although the variety of services the cell phones now offer has been vastly extended, the voice communication still occupies the position of a most important kind of communication.

Since both technologies are needed to provide constantly high quality of service in all areas, the voice communication cannot be excluded because of its leading position. To successfully measure and control the quality of service, or quality of speech to be more specific, in GSM/UMTS environment we can take advantage of multiple algorithms that have been invented for the IP based networks, which would allow us to create a flexible and low cost platform for GSM/UMTS speech quality measurement. By embracing this approach and creation of such a platform we would be allowed to measure influence of almost any signal interference from attenuation to high number of subscribers. The influence of the actual weather conditions on the speech quality in the GSM/UMTS networks is also measureable and it is a main topic of this paper.

In further chapters we will try to describe the measuring mechanism and the whole platform as well as the algorithm used for evaluation of the collected results and will also try to find whether there is a correlation between the speech quality and the current weather conditions.

Manuscript received March 8, 2012. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 218086 and Specific Research project No. 19742/2012 of Faculty Electrical Engineerging and Computer Science, VSB-Technical University of Ostrava.

Miroslav Voznak is an associate professor with the Department of Telecommunications, VSB –Technical University of Ostrava, Ostrava, Czech Republic (phone: +420 597321699; e-mail: miroslav.voznak@vsb.cz).

Jan Rozhon is a PhD. student with the Department of Telecommunications, VSB –Technical University of Ostrava, Ostrava, Czech Republic (phone: +420 597321641; e-mail: jan.rozhon@vsb.cz).

II. Speech Quality Assessment

Currently there are several techniques for speech quality measurement. Most of them have been developed for the VoIP, because of its natural susceptibility to speech quality degradation due to packet loss, jitter and other negative effects in IP networks [3]-[5].

A. Basic Scale and Rating

There are two main categories of speech quality assessment techniques – subjective and objective, the output of which is the Mean Opinion Score (MOS), which is a five degree scale for speech quality evaluation developed by ITU-T.

The basic scale as prescribed by the recommendation is depicted on Fig. 1. In order to avoid misunderstanding and incorrect interpretation of MOS values, ITU-T published ITU-T recommendation P.800.1 clearly describing Mean Opinion Score terminology [6]. This recommendation defines scales both for subjective and objective methods as well as for individual conversational and listening tests [7]-[9].

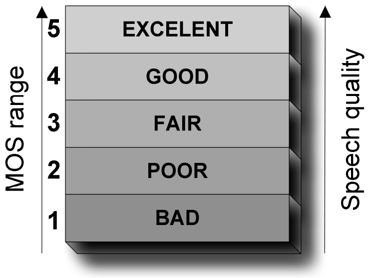


Fig. 1. MOS Scale.

The former subjective methods require a high number of actual listeners, who evaluate the quality of speech in absolute terms (ACR – Absolute Category Rating), where the actual comprehensibility and fidelity of a single voice sample is evaluated, or in the relative ones, where two voice samples are compared and the impairment between them is evaluated (DCR – Degradation Category Rating). The disadvantages of these methods come from the high time and cost ineffectiveness, since the high number of listeners needs to be involved and therefore the time required to perform the test is greatly increased. The objective methods try to eliminate these issues and introduce a human independent approach to speech quality. This can be achieved by substitution of humans in testing by mathematical computational models and algorithms. These methods use the MOS scale as well, or can output a number which then can be mapped on the MOS value using the defined conversion functions. The main goal of objective methods is as precise as possible estimation of the MOS value as it would be obtained by the subjective methods with the number of participants high enough to perform reasonable statistical analysis. Due to the effectiveness in the area of time consumption and economy, these methods are highly preferred and became the main mean to evaluate the speech quality in IP networks. This made them the very important topic of business and technological research, which resulted in the emergence of multiple algorithms that can be used [10]-[12]. To successfully understand this topic, which is of key importance to this paper, we need to distinguish two separate sub-groups in the objective methods – Intrusive and Nonitrusive.

B. Intrusive and Non-intrusive Approach

The intrusive methods work similarly to DCR. They use the original voice sample as it has entered the communication chain and the degraded one as it has been outputted by this communication chain. There are several algorithms that work this way, the following non-exhaustive list contains the most important ones.

- PSQM (Perceptual Speech Quality Measurement),
- PAMS (Perceptual Analysis Measurement System),
- PESQ (Perceptual Evaluation of Speech Quality),
- P.OLQA (Perceptual Objective Listening Quality Assessment).

The last mentioned algorithm, P.OLQA, is intended to be a successor of the PESQ and it tries to avoid the weaknesses of the PESQ's model and to incorporate the possibility of high bandwidth signal analysis.

Contrary to intrusive methods which need both the output (degraded) sample and the original sample, non-intrusive methods do not require the original sample. This is why they are more suitable to be applied in real time. Yet, since the original sample is not included, these methods frequently contain far more complex computation models. Examples of these types of measurements frequently use INMD (in-service nonintrusive measurement device) that has access to transmission channels and can collate objective information about calls in progress without disrupting them. These data are further processed using a particular method, with a MOS value as the output. The method defined by ITU-T recommendation P.563 or a more recent computation method E-model defined by ITU-T recommendation G.107 are examples of such measurements [11].

Today we can see various implementations of the speech quality testing mechanisms and algorithms mainly in business solutions. These solutions are conformant with the specifications of the International Telecommunication Union only from some part therefore the measurements cannot be compared without the thorough knowledge of the used algorithms.

On the other hand the telecommunication union itself presents on its websites the simple implementation of one of the most advanced algorithms in the field of speech quality measurement which we have mentioned. The PESQ algorithm is available for download in the form of source code and can be used to determine the conformance of the user developed solution with the ITU standard. Several open source programs are also built upon this algorithm and the companies (Optikom, Psytechnics), which developed the source code also offer their own services based on this source code and its modifications.

Regarding the speech quality in GSM and 3G networks mainly the first named company Optikom offers some services [13], but no one has performed the long term measurement with the focus on determination of weather influence on the speech quality in the GSM networks.

C. Perceptual Evaluation of Speech Quality

From the mentioned algorithms PESQ is currently the most common one. It combines the advantages of PAMS (robust temporal alignment techniques) and PSQM (exact sensual perception model) and is described in ITU-T's recommendation P.862 [8]. Since this algorithm is most widely used these days and since it generates results with reasonable accuracy and efficiency we have decided to use this algorithm as the keystone of our testing platform. The algorithms basic philosophy is depicted on the Fig. 2.

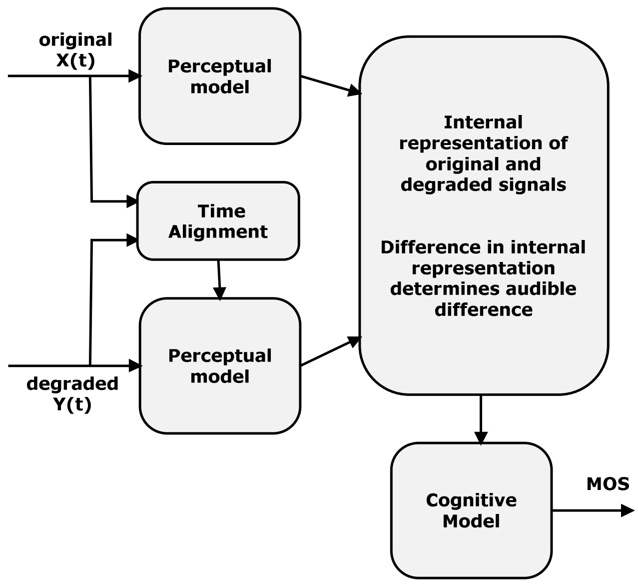


Fig. 2. Basic diagram of PESQ algorithm.

The best way of the speech quality assessment is the subjective approach. PESQ method was designed in order to provide as most accurate results, although correlations between objective and subjective scores in the benchmark were around 0.935 for both known and unknown data, the PESQ algorithm cannot be used to replace subjective testing.

It should also be noted that the PESQ algorithm does not provide a comprehensive evaluation of transmission quality. It only measures the effects of one-way speech distortion and noise on speech quality. The effects of loudness loss, delay, sidetone, echo, and other impairments related to two way interaction (e.g. centre clipper) are not reflected in the PESQ scores. Therefore, it is possible to have high PESQ scores, yet poor quality of the connection overall [8].

Most information on the performance of PESQ is from ACR listening quality (LQ) subjective experiments. This Recommendation should therefore be considered to relate primarily to the ACR LQ opinion scale.

PESQ compares an original signal X(t) with a degraded signal Y(t) that is the result of passing X(t) through a communications system. The output of PESQ is a prediction of the perceived quality that would be given to Y(t) by subjects in a subjective listening test. In the first step of PESQ a series of delays between original input and degraded output are computed, one for each time interval for which the delay is significantly different from the previous time interval. For each of these intervals a corresponding start and stop point is calculated. The alignment algorithm is based on the principle of comparing the confidence of having two delays in a certain time interval with the confidence of having a single delay for that interval. The algorithm can handle delay changes both during silences and during active speech parts. Based on the set of delays that are found PESQ compares the original (input) signal with the aligned degraded output of the device under test using a perceptual model. The key to this process is transformation of both the original and degraded signals to an internal representation that is analogous to the psychophysical representation of audio signals in the human auditory system, taking account of perceptual frequency (Bark) and loudness (Sone). This is achieved in several stages: time alignment, level alignment to a calibrated listening level, time frequency mapping, frequency warping, and compressive loudness scaling.

The internal representation is processed to take account of effects such as local gain variations and linear filtering that may – if they are not too severe – have little perceptual significance. This is achieved by limiting the amount of compensation and making the compensation lag behind the effect. Thus minor, steady state differences between original and degraded are compensated. More severe effects, or rapid variations, are only partially compensated so that a residual effect remains and contributes to the overall perceptual disturbance. This allows a small number of quality indicators to be used to model all subjective effects. In PESQ, two error parameters are computed in the cognitive model; these are combined to give an objective listening quality MOS. The basic ideas which are used in PESQ are described in [14].

Subjective votes are influenced by many factors such as the preferences of individual subjects and the context (the other conditions) of the experiment. Thus, a regression process is necessary before a direct comparison can be made. The regression must be monotonic so that information is preserved, and it is normally used to map the objective PESQ score onto the subjective score. A good objective quality measure should have a high correlation with many different subjective experiments if this regression is performed separately for each one, and in practice, with PESQ, the regression mapping is often almost linear, using a MOS like scale.

A preferred regression method for calculating the correlation between the PESQ score and subjective MOS, which was used in the validation of PESQ, uses a 3rd order polynomial constrained to be monotonic. In most cases, condition MOS is the chosen performance metric, so the regression should be performed between condition MOS and condition-averaged PESQ scores. A condition should at least use four different speech samples. The result of the regression is an objective MOS score in that test.

The closeness of the fit between PESQ and the subjective scores may be measured by calculating the correlation coefficient. Normally this is performed on condition averaged scores, after mapping the objective to the subjective scores. The correlation coefficient is calculated with Pearson's formula (1).

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \qquad (1)$$

In this formula, x_i is the condition MOS for condition i, and $\bar{x}$ is the average over the condition MOS values, $x_i \cdot y_i$ is the mapped condition-averaged PESQ score for condition i, and $\bar{y}$ is the average over the predicted condition MOS values y_i.

For 22 known ITU benchmark experiments, the average correlation was 0.935. For an agreed set of eight experiments used in the final validation – experiments that were unknown during the development of PESQ – the average correlation was also 0.935.

The human ear performs a time-frequency transformation. In PESQ this is implemented by a short-term FFT with a window size of 32 ms. The overlap between successive time windows (frames) is 50 percent. The power spectra – the sum of the squared real and squared imaginary parts of the complex FFT components – are stored in separate real valued arrays for the original and degraded signals.

The signed difference between the distorted and original loudness density is computed. When this difference is positive, components such as noise have been added. When this difference is negative, components have been omitted from the original signal. This difference array is called the raw disturbance density.

The final PESQ score is a linear combination of the average disturbance value and the average asymmetrical disturbance value. The range of the PESQ score is –0.5 to 4.5, although for most cases the output range will be a listening quality MOS-like score between 1.0 and 4.5, the normal range of MOS values found in an ACR experiment.

III. Measurement Platform

Our main goal was to create a testing platform that would be able to generate GSM calls automatically in regular intervals and together with that to log the actual weather conditions and analyze the voice sample in accordance to P.862 [9]. By using this platform we would be able to generate statistically significant amount of input data to perform the data mining analysis and find the possible correlation between one or multiple weather attributes and the obtained MOS value describing the quality of speech.

A. Collected Data

The form of the collected data has already been outlined. The server with Asterisk PBX has all the required information at its disposal thanks to the described algorithms and scripting mechanisms. The most important information includes:

- MOS value,
- Current temperature,
- Current humidity,
- Current dew point,
- Current rain,
- Current wind speed,
- Atmospheric pressure.

Before the server performs the insertion into the database table, it also performs MOS value normalization accordingly to what was said in the previous section. After this step the insertion of data into the database takes place creating a database table as it is outlined in a simplified way in the Tab. 1.

TABLE I
Simplified Version of the Database Table

id	Date	Hour	Minute	MOS	Meteorol. Data
1	2011-10-16	16	5	3.456	Data
2	2011-10-16	16	10	3.398	Data
3	2011-10-16	16	15	3.361	Data

B. Implementation of Measurement Platform

To successfully achieve this goal, we decided to use a Linux based system due to its natural effectiveness in the field of automation and process monitoring. Moreover this would allow the whole platform to be highly cost-effective.

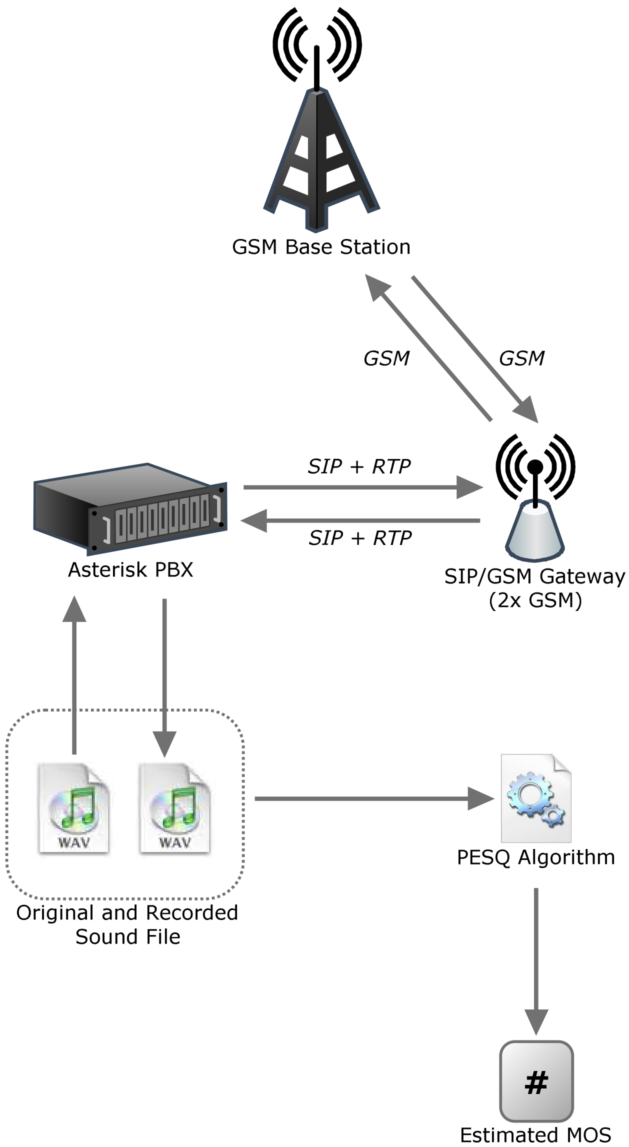


Fig. 1. Scheme of testing platform with the detail on MOS computation.

On the Linux machine, we configured the most commonly used VoIP PBX Asterisk in the way that it was able to generate one call every five minutes over the SIP (Session Initiation Protocol) and RTP (Real-time Transport Protocol). These calls were routed to a SIP peer, which represented a SIP/GSM+UMTS gateway with two separate SIM cards. Each call was routed from the SIM card in module one, to the SIM in module two, thus allowing to use single BTS station in the building nearby and consequently minimize the interference caused by the BTS switching or long signal routes to best possible minimum since the measured is being performed on the closest BTS. The GSM gateway then routed the incoming calls back to Asterisk PBX, which recorded the voice data to a separate WAV file. This way the loop was created allowing the Linux system with Asterisk PBX to have access to both original WAV file and the degraded one, which is necessary for successful implementation of PESQ algorithm. The hardware requirements were low and are listed below:

- Low-end HW server with Ethernet interface,
- SIP/GSM+UMTS gateway with two separate modules working on 1800MHz (DCM-1800).

As the speech files we used the samples of calibrated voices specifically designed for the use with PESQ algorithm with the sampling frequency of 8 000 Hz, encoded using PCM into 128 kbps stream of one channel audio. The cooperation between Asterisk PBX and PESQ algorithm was achieved via the "System" command in Asterisk dialplan that fires the python script, which performs degraded file modification, fetching the current meteorological data, performing the PESQ evaluation and storing the data into the database.

The degraded file modification was made necessary due to about 2 seconds of leading period in the degraded file, where two ring tones from Asterisk were recorded. If this was left unhandled, the results from the PESQ analysis would be highly inaccurate and therefore not useful.

The meteorological data were obtained from the local meteorological station working in the university campus about 300 meters from the BTS station. The actual data transmission was implemented by HTTP communication, where python script asked meteorological station for current conditions in text format, which then was parsed and data were stored.

As a database the SQLite engine is used because of two reasons. Firstly, the configuration and manipulation of the database are quite simple in this database engine but still providing enough features to successfully complete the given task without any complication caused by the engine limitations. And secondly, the whole database can be backed up easily [4]. This is allowed by the fact, that the whole database is stored one user-defined file. Especially the latter is important due to the long period of measurement.

Storing data into the database is useful also from the data analysis point of view, because the values of measured MOS can be displayed in almost real time with the use of the conditional database lookup for any month, day or even hour or minute thanks to the used timestamp. Moreover the aggregate functions allow quick data analysis even on the huge number of stored values.

The software equipment used to perform the measurement consists of:

- Ubutnu 10.04.1 x64,
- Asterisk PBX 1.6.2.16,
- SQLite3.

The testing platform scheme is depicted on the Fig.2, where the basic communication and protocol are displayed in the clearer detail.

C. Basic Assumptions

From the previously presented architecture it is clear that the GSM network and the parameters affecting it is not the only factor that contributes to the final speech quality. In addition the status of the network over which Asterisk PBX and SIP/GSM gateway communicate can influence the speech quality. Moreover the codec translation between used codecs (G.711A, GSM) can have deteriorating effect on the speech quality devaluing the final results. Because these inherent effects cannot be eradicated basic assumptions applied to measurement had to be designed.

Firstly the communication between Asterisk PBX and SIP/GSM gateway takes place on the 100Mbps Ethernet line, where no other traffic is allowed. Therefore the network will never reach the congestion state and all the information on the network is exchanged as quickly as it is physically possible for this type of interconnection. Because of this the delay of packets and its variation is minimal and does not need to be counted with.

The codec translation is not a parameter, which can be easily dealt with. Since the SIP/GSM gateway supports only two codes for the VoIP communication (namely G.711 and G.729) and the GSM communication is built upon the GSM EFR codec the translation will take place every time the call is made.

Therefore the measured MOS value will always be affected by this process. To eliminate or at least diminish the influence to speech quality we do a calibration of the system. This process takes the ideal MOS value of speech quality using the GSM EFR codec as the basis, to which we compare the best result taken from several hundred measurements. The difference between the ideal and best case result is then identified as the distortion caused by the codec translation and this value is then added to all the measured values.

This way the codec translation influence is limited and the measurement will provide reasonable results. Even if this countermeasure was not performed the trend in the MOS values would still be preserved and the correlation between weather condition and measured values could still be found.

Other approach could be built upon creating the reference codec translation result in the exclusively IP based network, but as it has been explained in the previous text, this step could be skipped and the reasonable data still can be obtained.

D. *Data Analysis*

As the convenient and efficient way of data analysis clustering algorithms allow for standardized and confirmed data segregation into the groups with similar attributes. The simplest definition is shared among all and includes one fundamental concept: the grouping together of similar data items into clusters. These obtained clusters should reflect some mechanism at work in the domain from which instances or data points are drawn, a mechanism that causes some instances to bear a stronger resemblance to one another than they do to the remaining instances.

Let $X \in R^{mxn}$ a set of data items representing a set of m points x_i in R_n. The goal is to partition X into K groups C_k such every data that belong to the same group are more "alike" than data in different groups. Each of the K groups is called a cluster. The result of the algorithm is an injective mapping $X \rightarrow C$ of data items X_i to clusters C_k [17].

Since there are multiple clustering methods and algorithms, the efficient and most precise one for our purposes needed to be found empirically. Through several testing runs of several algorithms (K-means clustering, EM clustering…) K-means served best in our case meaning that the clusters incorporated the logically most correct data and did not suffer from the algorithm's tendency to create the clusters of equal or similar size. The definition of K-means clustering looks as follows. Let $S = \{X_1, X_2, ..., X_N\}$ be a dataset with n observations, each of which is p-dimensional. The objective in K-means clustering is to group these observations into categories C_1, C_2,…,C_K for given K, such that the objective function is minimized (2).

$$O_K = \sum_{i=1}^{n} \sum_{k=1}^{K} I(X_i \in C_k)(X_i - \mu_k)^{'}(X_i - \mu_k) \qquad (2)$$

Here μ_k represents the mean vector of observations from C_k,

$$\mu_k = \frac{1}{n_k} \sum_{i \in C_k} X_i \qquad (3)$$

where $n_k = |C_k|$ is the number of observations in C_k and $I(X_i \in C_k)$ is an indicator function specifying whether observation X belongs to the k-th group. Further, note that the following $\|x\| = \sqrt{x'x}$ denotes the Euclidian norm of p-dimensional vector x.

IV. Results

As we put the pieces together we can finally analyze the results. Using the mentioned K-means clustering method we have performed analysis of all available possible influencer of MOS value in GSM/UMTS environment. To be more specific, we explored the influence of the Current Temperature, Humidity, Rain, Dew Point, Wind Speed and Atmospheric Pressure.

In addition we had even other parameters at our disposal, such as THW Index, Wind Direction and so on, but these are expected to have no impact on the measured MOS value because of their signal non-interfering nature.

Through series of data mining operation we came to conclusion, that the self-correcting mechanisms implemented in the GSM/UMTS technology prevent call from being interfered by weather condition. The fact that there is no statistically significant relation between MOS and Humidity, which is the most influencing factor from the signal strength point of view, discouraged us from the thought that there might be a relation of any kind. Because of this we didn't expect that the other parameters connected directly to humidity like current rain rate would be influencing the voice quality with a measurable significance. However, this correlation can be found in the statistical data.

The Tab. 2 shows the actual probability of MOS value depending on the current rain density. Second row in the table can be seen as the most important and tells us the following: if the current rain density is between 28.5 and 33.9 mm/h we can expect MOS to be lower than 2.599 with the probability of 64%. If the rain density gets over 33.899 mm/h than we can expect MOS lower than 2.599 with 100% probability. Other rows can be read similarly. These results in the following, with the increasing rain activity, the MOS value drops significantly. Especially when the high rain density is reached (greater than 5 mm/h) the MOS value drops to level, where the user can experience very bad speech quality and low comprehensibility. The rain density up to 5 mm/h practically does not affect speech quality.The last row is actually a conjunction of two separate clusters, which were identified as distinct areas by the algorithm. However they are mutually complementary and therefore they were combined into single row.

TABLE II.
THE INFLUENCE OF CURRENT RAIN DENSITY ON MOS.

Rain Rate [mm/h]	Favors	Relative Impact
>= 33.899	< 2,599	100
28.502 - 33.899	< 2,599	64
15.133 - 28.502	2,599 – 2,908	100
4.883 – 15.133	2,908 – 3,281	100
< 4.883	>=3,281	100

If we put that together with our knowledge of low humidity influence, we can state that this quality drop occurs in the beginning of the rain, while the air humidity is still low and the effect diminishes while the humidity rises. This can be caused by the slow adaptation of the network or mobile station to the rain.

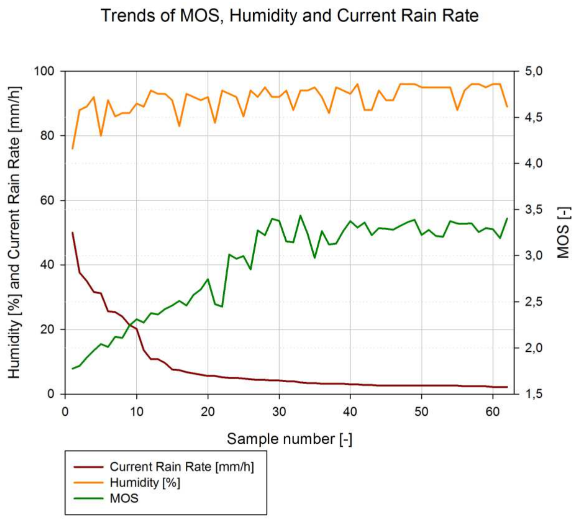


Fig. 2. Relation between MOS, Humidity and Current Rain Rate.

The data of high significance (with rain density) can be interpreted as the sequence resulting in the chart on the Fig.3. Here we can see the data subset containing the samples with the significant rain density (60 samples with rain density higher than 2.2 mm per hour) ordered from the highest to the lowest rain density. As we can observe the humidity varies independently on the current rain density, however MOS factor is influenced measurably.

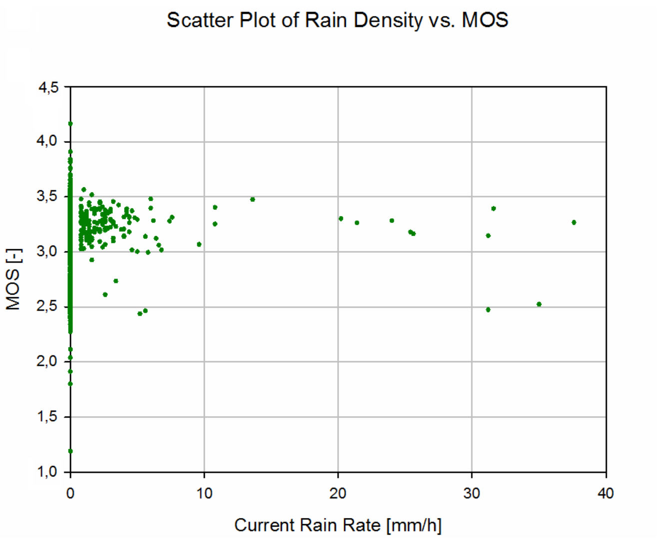


Fig. 4. Scatter Plot of Rain Density vs. MOS.

Fig. 3 clearly confirms the stated information about the rain influence on the quality of speech in GSM/UMTS environments. As already stated this is the only significant relation, therefore no further analysis makes sense at this point. To complete the picture, it is necessary to state that the measured data were collected from August to November of 2011 and the total data pool contains about 20 000 data rows. Since the BTS station and all the other parts of the network, through which the signal traverse (MSC, IMS Core), are not under our control and therefore the information about their load is not possible to obtain, the statistical methods were used to eliminate the effect of variable load of these network elements using huge amount of input data.

The scatter plot depicted in Fig. 4 illustrates the relation between MOS and Current Rain Rate [mm/h].

V. Conclusion

Through long term measurement we have obtained both meteorological data and MOS values specifying the current speech quality in the GSM networks. By utilization of advanced techniques in data mining and data analysis, our team found the correlation between current rain density and the speech quality. This shows us a great decrease of speech quality in the earliest phases of the rain, where the humidity is still low. Particularly, we can observe nearly a 50 percent decrease in measured MOS parameter when comparing sample results obtained during the heavy rain and those obtained during a mild shower. This decrease is reflected in the worsened speech quality by glitches in the speech, low comprehensibility and other communication difficulties. The reason for this behavior can be found in the BTS transmitted power adjustment, which can modify the level in a range of up to 20dB [15,16].

The slow response of the correction procedure in the BTS can cause problems with the same nature as we have witnessed during our measurement. Other possible explanation includes the influence of the number of subscribers logged in the particular BTS at a particular time, but this would result in quality deterioration mainly during the busy hour and this behavior was not observed. No other significant bond was found.

By performing this measurement, we have proved that the low cost measuring platform can be developed and used for the speech quality measurement in the GSM networks. We have successfully taken advantage of our team’s knowledge in the IP telephony and transited this knowledge to cellular networks.

The greatest possible improvement of our method as we see it now is to perform the measurement during the whole year to have complete knowledge of the speech quality trends during all the possible weather conditions and to modify the testing platform by introducing OpenBTS solution, which would allow for gaining the full control over the transmission chain.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 218086 and Specific Research project No. 19742/2012 of Faculty Electrical Engineerging and Computer Science, VSB-Technical University of Ostrava.

REFERENCES

[1] M. Kavacky, E. Chromy, L. Krulikovska and P. Pavlovic, "Quality of Service Issues for Multiservice IP Networks, " *In Proc. SIGMAP 2009 International Conference on Signal Processing and Multimedia Applications*, Milan, Italy, July 2009, pp. 185 – 188.

[2] A. Maraj and S. Rugova, "The role of softswitch in NGN network," In Proc. *8th WSEAS International Conference on Data Networks, Communications, Computers*, November 2009, Baltimore, pp. 179-184.

[3] F. Alvarez-Vaquero and J. Sanz-Gonzalez,"Network VoIP for corporative environment design," In Proc. *7th WSEAS International Conference on Telecommunications and Informatics*, May 2008, Istanbul, Turkey, pp. 194-198.

[4] I. Baronak and M. Halas, "Mathematical representation of VoIP connection delay," *RADIOENGINEERING* Volume: 16 Issue: 3, September 2007, pp. 77-85.

[5] A. Kovac, M. Halas, M. Orgon, M. Voznak, "E-model MOS Estimate Improvement through Jitter Buffer Packet Loss Modelling," In *Advances in Electrical and Electronic Engineering*, Volume 9, Number 5, December 2011, pp. 233-242.

[6] ITU-T P.800.1 - Mean Opinion Score (MOS) terminology, *ITU-T*, July 2006.

[7] ITU-T P.800 – Recommendation P.800 of the International Telecommunication Union, Methods for subjective determination of transmission quality, *ITU-T*, 1996

[8] ITU-T P.862 – Perceptual evaluation of speech quality (PESQ): An objective Method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, *ITU-T*, February 2001

[9] J. Rozhon, M. Voznak, "Development of a speech quality monitoring tool based on ITU-T P.862," *34th International Conference on Telecommunications and Signal Processing, TSP 2011 - Proceedings*, Budapest, August 2011, pp. 62-66

[10] M. Voznak, M. Tomes, Z. Vaclavikova, M. Halas, "E-model improvement for speech quality evaluation including codecs tandeming," *International Conference on Data Networks, Communications, Computers - Proceedings*, Faro, November 2010, pp. 119-124

[11] ITU-T Recommendation G.107, The E-model: A computational model for use in transmis-sion planning, *ITU-T*, Geneva, 04/2009.

[12] M. Voznak, "E-model modification for case of cascade codecs arrangement," *International Journal of Mathematical Models and Methods in Applied Sciences*, Volume 5, Issue 8, 2011, pp. 1439-1447.

[13] Opticom: Opticom quality testing products page [Online]. Available: http://www.opticom.de/products/product-overview.php.

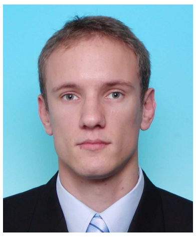
[14] A. Rix, R. Reynolds, M. Hollier, "Perceptual measurement of end-to-end speech quality over audio and packet-based networks," 106-th *AES Convention*, No. 4873, May 1999.

[15] M. Mouly, M. Pautet, "The GSM System for Mobile Communications," *Telecom Pub-lishing*, 1992.

[16] S. Saunders, A. Zavala, "Antennas and Propagation for Wireless Systems," *Wiley*, 2nd edition, 2007.

[17] G. Fung, "A Comprehensive Overview of Basic Clustering Algorithms," University of Wisconsin-Madison, 2001 [Online]. Available: http://pages.cs.wisc.edu/~gfung/

Miroslav Voznak holds position as an associate professor with Department of Telecommunications, Technical University of Ostrava. He received his M.S. and Ph.D. degrees in telecommunications, dissertation thesis "Voice traffic optimization with regard to speech quality in network with VoIP technology" from the Technical University of Ostrava, in 1995 and 2002, respectively. Topics of his research interests are Next Generation Networks, IP telephony, speech quality and network security.



Jan Rozhon received his M.S. degree in telecommunications from VSB – Technical University of Ostrava, Czech Republic, in 2010 and he continues in studying Ph.D. degree at the same university. His research is focused on performance testing of NGN and in this field he cooperates with CESNET association. He received rector's appreciation for scientific contribution of his diploma thesis in 2010.