Communications and Network, **2013, 5, 448-454**
http://dx.doi.org/10.4236/cn.2013.53B2083 Published Online September 2013 (http://www.scirp.org/journal/cn)

Radio Link Parameters Based QoE Measurement of Voice Service in GSM Network*

Wenzhi Li[1], Jing Wang[1], Zesong Fei[1], Yuqiao Ren[1], Xiao Yang[2], Xiaoqi Wang[2]
[1]School of Information and Electronics, Beijing Institute of Technology, Beijing, China
[2]The Research Institution of China Mobile, Beijing, China
Email: wenzhi306@163.com, wangjing@bit.edu.cn, feizesong@bit.edu.cn, ryq8884291@126.com,
yangxiao@chinamobile.com, wangxiaoqiyf@chinamobile.com

Received May 2013

ABSTRACT

Recently, Quality of Experience (QoE) of voice service has been paid more attentions because it represents the performance of voice service subjectively perceived by the end users. And speech quality is commonly used to measure the QoE value. In this paper, a speech quality assessment algorithm is proposed for GSM network, aiming to predict and monitor QoE of voice service based on radio link parameters with low complexity for operators. Multiple Linear Regression (MLR) and Principal Component Analysis (PCA) are combined and used to establish the mapping model from radio link parameters to speech quality. Data set for model training and testing is obtained from real commercial network of China Mobile. The experimental results show that with sufficient training data, this algorithm can predict radio speech quality with high accuracy and could be used to monitor speech quality of mobile network in real time.

Keywords: QoE; Speech Quality Assessment; Voice Service; Regression Analysis; PCA; GSM

1. Introduction

Voice service has been and will continue to be the most fundamental and significant service in cellular mobile communication systems. And speech delivered over Global System for Mobile Communications (GSM) network accounts for much of voice traffic. Therefore, for operators, it is of significant that Quality of Experience (QoE) [1] of voice service can be monitored in real time, which guides network optimization as well as network maintenance directly and effectively. Speech quality is considered as the most comprehensive metric that characterizes the QoE of end subscriber. A QoE measurement algorithm, which can reflect the radio link condition and could be integrated in the signaling monitor system, is preferred from the perspective of operators. Note that the novel algorithm should be real-time and accurate. Besides, low complexity is also necessary.

Subjective Mean Opinion Score (MOS) [2] assessment reflects the listener's actual perception of voice best, but the operation is time-consuming and laborious. Thus, objective assessments algorithm, which can be divided into voice based and radio link parameters based algorithms depending on whether voice signal is needed, is developed to approximate the subjective MOS. Perceptual Evaluation of Speech Quality (PESQ) [3] proposed by ITU-T is a typical voice-based algorithm which is a commonly used method of voice quality test in wireless network due to its quite high relevance with subjective MOS. However, PESQ does not apply to the long-term and large scale network monitoring for its high cost of implementation. The assessment method based on network parameters [4] is more suitable to real-time assessment of voice quality in mobile network, because most of its input parameters can be measured from network in real time.

The Speech Quality Indicator (SQI) [5] algorithm developed by Ericsson Corporation and the Voice Quality Index (VQI) [6] specifically for Time Division Synchronous Code Division Multiple Access (TD-SCDMA) network by Huawei Corporation are two typical algorithms based on network parameters. SQI expresses the degree of voice distortion caused by radio link transmission, which is calculated by weighting a number of radio link parameters including Bit Error Rate (BER), Frame Error Rate (FER), handing over, Discontinuous Transmission (DTX) and speech coding mode (speech codec), etc. VQI has a similar thinking with SQI. The input parameters of VQI are speech coding mode, FER, BER, handing over and frame loss.

*This work was supported by China National S&T Major Project (2012ZX03001034).

Although those two algorithms have been implemented by equipment manufacturers, with high accuracy if adequate enough network parameters are collected, their index values are not very applicable to monitor the QoE of voice service and network quality. That's because the major parameters such as FER, BER, frame loss and so on which have a great impact on QoE can't be real-time acquired by operators in the GSM signaling monitoring platform. Besides, the speech index values of SQI and VQI cannot be compared in the network monitoring and optimization because of their private interfaces by different manufacturers. The purpose of this paper is to solve the existing problems by proposing a novel QoE measuring algorithm especially for GSM network. The algorithm inputs are specific network parameters collected in signaling monitoring platform from commercial GSM network of China Mobile. Multiple Linear Regression (MLR) based on least squares is adopted to further investigate the relationship between network parameters and QoE of voice service. All of these features make it possible that the real-time algorithm with low complexity is suitable for monitoring QoE of voice service by operators.

2. Measurement of QoE of Voice Service Based on GSM Network Parameters

2.1. Thinking of Measurement Algorithm

The purpose of this algorithm is to measure QoE of voice service in real time by GSM network parameters. Therefore, two conditions should be satisfied: network parameters must be obtained in real time; a mapping model from network parameters to speech quality should be established.

In GSM network, Measurement Report (MR) is one of the main foundations to assess the quality of radio environment. The MR signaling is transmitted every 480ms in traffic channel (470ms in signaling channel), including Received Signal Quality (RxQual), Received Signal Level (RxLev), handing over, hopping, speech coding mode and etc. Therefore, selecting MR as the access of network parameters can not only express the quality of current radio link, but also requires little cost to transform current network. Considering the following conditions: time for human ear to percept voice, PESQ algorithm proposing the assessed object includes at least 3.2 s speech [7] and the quantity of MR demanded by measuring algorithm and the efficiency of data collection in commercial network, the time granularity of measuring algorithm used in this paper is set as 4.8 s finally.

The next step is to obtain the speech quality used for data modeling corresponding to network parameters. The specific approach is to record the voice sample corresponding to a set of network parameters in time, and then assess the speech quality with PESQ algorithm. The model mapping from network parameters to voice quality adopts the Multiple Linear Regression method which takes the advantage of low complexity and high accuracy.

2.2. Obtaining Network Parameters

To reflect the status of current network more realistically, both the model training and testing use data are collected from the commercial network. In order to accurately measure the influence to speech quality caused by radio link parameters, we captured the network parameters and speech data using the way of cell phone calls landline.

A communication circuit includes wired links and wireless links, of which the wireless links are the key aspects that affect speech quality while the wired parts having less effects on speech quality are negligible. Meanwhile, the algorithm is to assess the voice quality of one single side of the radio links because most of the parameters reflecting the network quality could not be transmitted to the other side of core network (MSC). Accordingly, the method of obtaining network parameters and speech samples is expressed in **Figure 1**.

The transmissions of uplink speech and downlink speech are relatively independent process. Therefore, the speech quality of uplink or downlink is affected by uplink or downlink respectively. The uplink parameters are measured by the Base Transceiver Station (BTS) of the network while the downlink parameters are measured by the user terminal and then reported to network by Measurement Report signaling of Um interface. In summary, both the uplink and down link parameters are collected by signaling monitoring platform of Base Station Controller (BSC).

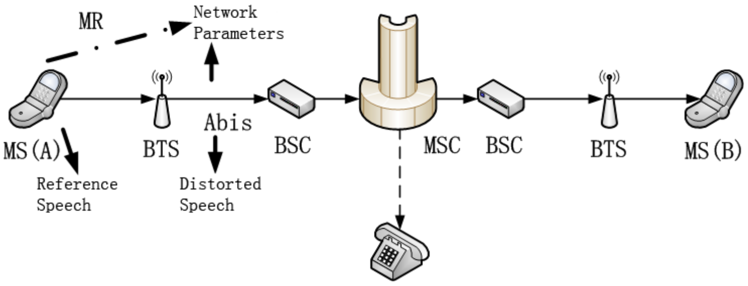

Figure 1. Model of data collection.

The uplink distorted speech should be obtained from BTS or BSC in theory, but it is not supported in real network and of more cost in transformation. Furthermore, considering the little loss of speech quality caused by wired transmission, we use the way that MS calls landline, and collect the distortion speech from landline side.

PESQ algorithm recommended by ITU P.862.1 is used to calculate the MOS value of every single speech because PESQ is the most widely used algorithm to assess speech quality in mobile network testing currently, which is of very high relevance with subjective MOS.

For every MR the absolute time was record accurately and for every voice sample the recording start time was recorded, in order to match the voice sample with MR conveniently. Specifically, every distorted voice was corresponded to 10 pieces of MR (480 ms) data, which was used for the training and verifying of algorithm.

2.3. The Specific Structure of the Algorithm

The structure of the algorithm is shown in **Figure 2**, with detailed description of each part shown as following.

2.3.1. Preprocessing the Data

The speech quality level for a certain period is related with not only the average level but also the fluctuation of the network parameters. To reflect the fluctuation of the network parameters, we calculated the mean, variance, extreme value and some other statistics of the 10 observations during 4.8 s. Specifically, we assume that the observation matrixes are Equations (1) and (2):

$$\begin{pmatrix} Rxq_{i1} & RxL_{i1} \\ \vdots & \vdots \\ Rxq_{i10} & RxL_{i10} \end{pmatrix} \tag{1}$$

$$\begin{bmatrix} codec_i & HO_i & HOP_i & DTX_i \end{bmatrix} \tag{2}$$

where i ranging from 1 to n refers to the speech sample index, n denotes the total number of observations, the Rxq_{ij} and RxL_{ij} stand for *RxQual* and *RxLevel* separately, and the *codec*, *HO*, *HOP* and *DTX* stand for speech coding mode, handing over happened or not, hopping used or not, discontinuous transmission used or not.

The first matrix was preprocessed, and the output is Equation (3) combined with Equation (2), where X_{ij} is the statistics of *RxQual* and *RxLevel*.

$$\begin{bmatrix} X_{i1} & \dots & X_{im} & codec_i & HO_i & HOP_i & DTX_i \end{bmatrix} \tag{3}$$

2.3.2. Data Classification

Collected data should be classified according to coding mode, because the network parameters influence the quality of speech transmission by different mechanism under different coding mode. Specifically, the total data was divided according to codec. Assuming the number of data collected under a certain coding mode (e.g. FER) is n, then the observed data matrix of this mode is:

$$\begin{bmatrix} X_{11} & \dots & X_{1m} & codec_1 & HO_1 & HOP_1 & DTX_1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n1} & \dots & X_{nm} & codec_n & HO_n & HOP_n & DTX_n \end{bmatrix} \tag{4}$$

2.3.3. Principle Components Extraction

The data have a larger dimension after preprocess, which make the analysis of relationship between preprocessed data difficult in multidimensional space. Besides, the parameters have very strong correlations with each other, which lead to cross impacts on the speech quality, and it will be difficult to analyze and present this cross effect. Principal component analysis was introduced to solve the problem. Specifically, we analyzed the correlation of the first m columns in the matrix expressed in section 2.3.2, using Principal Component Analysis (PCA) to calculate the principle components, and then we took the first p principle components of larger variance as the input vectors of regression analysis, that is:

$$\begin{bmatrix} Y_{11} & \dots & Y_{1p} \\ \vdots & \ddots & \vdots \\ Y_{n1} & \dots & Y_{np} \end{bmatrix} \tag{5}$$

in which every data column is a selected principle component.

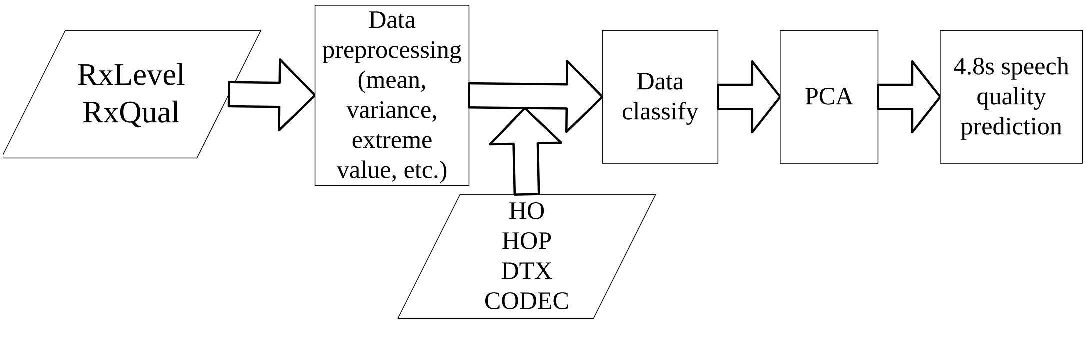
RxLevel
RxQual
Data preprocessing (mean, variance, extreme value, etc.)
HO
HOP
DTX
CODEC
Data classify
PCA
4.8s speech quality prediction

Figure 2. Basic structure of algorithm.

It should be noted that not all selected *p* principle components will necessarily remain in the final measuring formula, because some principle components of little impact on speech quality would be excluded according to the hypothesis testing results in the fitting procedure of regressing equation.

2.3.4. Quality Measuring of 4.8 s Speech

Under a certain coding mode (e.g. uplink FER), the basic form of the measuring formula is:

$$\begin{aligned} MOS_{upEFR} &= a_0 + a_1 * Y_1 + \cdots + a_i * Y_i + a_{i+1} * HO \\ &+ a_{i+2} * HOP + a_{i+3} * DTX \end{aligned} \tag{6}$$

where $Y_1 \cdots Y_i$ are the extracted principle components, and *HO*, *HOP*, *DTX*, *codec* are limited to specific discrete values, and a_i are fitted coefficients.

The final form of the measuring formula and the fitting coefficients would be obtained through multiple regression analysis [8]. For each coding mode, the preliminary least-squares fitting of the input data and the speech quality values would be taken, and then the test of significance (e.g. the F-test and T-test) will apply to the obtained regression equation. An F-test ($a = 0.05$) is used to determine whether the liner relationship of the equation is significant, while a T-test is used to determine whether the impact of each variable is significant, leading to some variables excluded according to the result. After the hypothesis testing, the regression equation needs residual test and outlier test to determine whether the nonlinear transform processing or some other kinds of processing should be taken to the data. Normality test is used in the residual analysis.

3. Performance Analysis

3.1. Condition of Data Collection

The data collection for algorithm training and testing was based on three typical codec modes of GSM network, with the network parameters and speech samples recorded according to uplink and downlink separately. The distribution of valid data used in the algorithm is shown as **Table 1**.

For each case, three quarters of the total data are used for algorithm training to produce the measuring formula of speech quality. The left one quarter data are used to test the performance of algorithm.

3.2. Evaluation Index of the Algorithm

Here, we call the QoE value of voice service predicted by radio link parameters in mobile network as RSQ (Radio Speech Quality). For each set of network parameters (corresponded to 10 pieces of MR data), we predict a RSQ value using this algorithm, and compare it with the actual PESQ value of the speech, counting the following indicators to measure the algorithm’s prediction accuracy. Aiming at monitoring speech quality in actual network, this paper proposed a stricter segmented relative error indicator.

- Indicator 1: Segmented relative error, as shown in **Table 2**. In order to eliminate the influence to statistic result caused by the interval endpoint value, normalized relative error expressed in Equation (7) is used based on the fact that PESQ (MOS-LQO) [9] has a working range of (1.02, 4.56].

$$relative\ error = \frac{\left| Actual\ MOS - Predicted\ MOS \right|}{\Delta} * 100\% \tag{7}$$

where $\Delta = \left(MOS_H - MOS_L \right) = 4.56 - 1.02$, MOS_H and MOS_L stand for upper and lower limits of the PESQ value.

Specially, in MOS range of (1.02, 2], the accuracy of low value alarm was taken to indicate the accuracy of the algorithm. That is because the referenced PESQ algorithm

Table 1. Distribution of data used in algorithm.

Actual PESQ value range	Uplink			Downlink		
	EFR	FR	HR	EFR	FR	HR
(1, 2]	82	20	4	13	20	2
(2, 3]	62	63	18	110	63	10
(3, 4.5]	413	251	185	326	251	198

Table 2. Relative error indicators in segments.

Actual RSQ	Output of algorithm	Segmented indicators of accuracy
(1.02, 2]	RSQ value and low value alarm (give low value alarm when predicted speech quality is in (1.02, 2])	Accuracy of low value alarm
(2, 3]	RSQ value	Percentage of data whose relative error is less than 10%
(3, 4.56]	RSQ value	Percentage of data whose relative error is less than 10%

itself has a low measuring accuracy in low value interval, and the speech quality become intolerable when MOS value is lower than 2, where it moots to give the specific MOS value, so alarm should be given when the actual network quality appears very low. The accuracy of low value alarm is calculated in Equation (8),

$$Accuracy\ of\ \ low\ value\ alarm=\frac{M}{N}. \qquad (8)$$

where *M* denotes the number of samples with predicted MOS in $\left(1.02, 2+0.2\right]$ and actual MOS in $\left(1.02, 2.0\right]$, *N* is the total number of samples with actual MOS in $\left(1.02, 2.0\right]$.

- Indicator 2: Pearson's correlation coefficient *R* calculated by Equation (9),

$$R=\frac{\sum_{i=1}^{N}\left(q_i-\overline{q}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(q_i-\overline{q}\right)^2\sum_{i=1}^{N}\left(y_i-\overline{y}\right)^2}} \qquad (9)$$

where q_i and $\overline{q}$ stand for the value and mean of actual MOS separately; y_i and $\overline{y}$ stand for the value and mean of predicted MOS by algorithm separately.

- Indicator 3: Root Mean Square Error *RMSE* is shown in Equation (10), where q_i and y_i stand for the actual MOS value and the predicted MOS value separately.

$$RMSE=\sqrt{\frac{\sum_{i=1}^{N}\left(q_i-y_i\right)^2}{N}} \qquad (10)$$

Correlation coefficient and Root Mean Square Error are metrics of performance commonly used in international objective quality assessment algorithm, which can measure not only the correlation between the predicted value and the real value but also the degree of dispersion.

3.3. Test Performance of the Algorithm

The valid data collected were divided into training data accounted for three quarters and testing data accounted for one quarter. Accuracy in indicator 1 (**Table 2**) is shown as **Table 3** ("-" indicates amount of data of the interval is too small to count accuracy).

Due to the network conditions, the data collected from commercial network was difficult to achieve traversal, and most networks in the collection area were configured EFR mode to achieve relatively better performance, with fewer FR and HR data. Accordingly, the algorithm has a better performance under EFR mode because of more training data. **Table 4** shows the measuring results of indicator *R* and *RMSE* under EFR mode.

To reflect the performance of algorithm intuitively, the maps of actual RSQ and predicted RSQ under uplink EFR mode is shown for example in **Figure 3**.

Table 3. Accuracy of algorithm in indicator 1.

Actual PESQ value range		(1, 2]	(2, 3]	(3, 4.5]
Uplink EFR	training effect	93%	60%	96%
	testing effect	95%	58%	94%
Downlink EFR	training effect	67%	65%	92%
	testing effect	100%	67%	93%
Uplink FR	training effect	71%	54%	88%
	testing effect	100%	56%	88%
Downlink FR	training effect	-	83%	99%
	testing effect	-	100%	97%
Uplink HR	training effect	67%	73%	98%
	testing effect	-	-	98%
Downlink HR	training effect	-	85%	97%
	testing effect	-	67%	95%

Table 4. *R* and *RMSE* in EFR mode.

Actual PESQ value range		(1, 2]	(2, 3]	(3, 4.5]	Overall
Uplink EFR training results	*R*	86%	91%	84%	97%
	RMSE	0.19	0.27	0.17	0.19
Uplink EFR testing results	*R*	86%	82%	83%	97%
	RMSE	0.20	0.25	0.17	0.19
Downlink EFR training results	*R*	66%	52%	89%	92%
	RMSE	0.50	0.28	0.19	0.23
Downlink EFR testing results	*R*	82%	56%	89%	91%
	RMSE	0.50	0.29	0.21	0.25

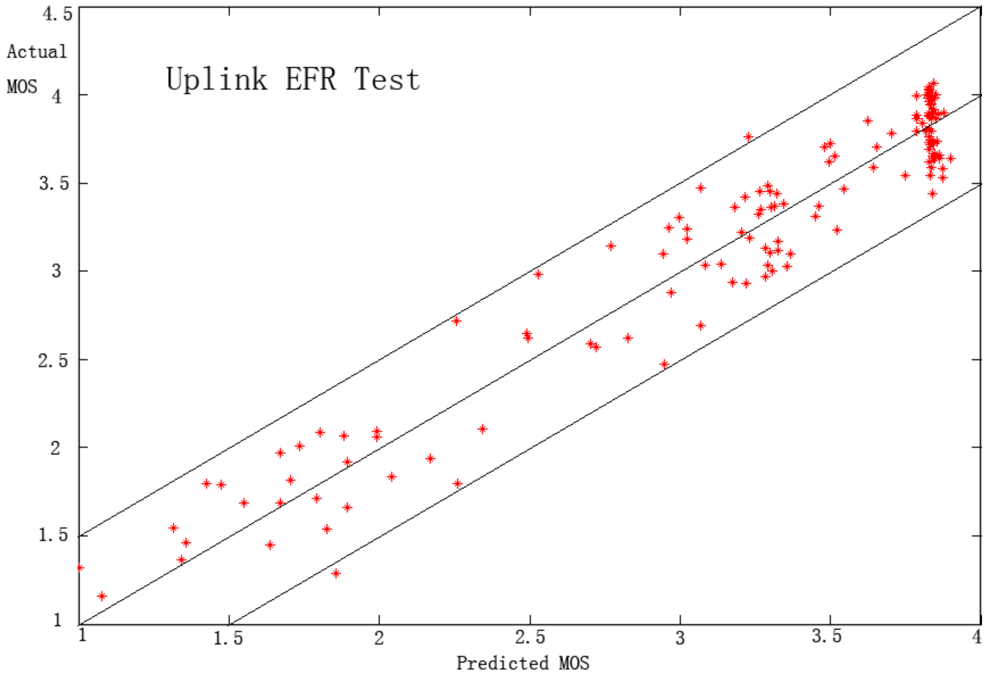

Figure 3. Distribution of Actual MOS and Predicted MOS for UP-EFR Mode.

In **Figure 3**, the abscissa indicates the RSQ predicted by the algorithm using radio link parameters, and the ordinate indicates the speech quality assessed by PESQ. The middle is the 45˚ isoline, on which the predicted values and the actual values are equal. And two lines which indicate that the absolute value of estimated error is 0.5 are below and above the isoline.

It can be suggested, for uplink EFR, the amount of data is adequate and the MOS values distribute relatively evenly, which means the amount and ergodicity are better. Consequently, the overall correlation is greater than 90%, and the *RMSE* is about 0.2, indicating that the measuring algorithm is of better performance, In addition, accuracy of training and testing are basically equal, showing a good stability of proposed algorithm. Meanwhile, amount of data is much larger in MOS interval of (3, 4.5] than (2, 3], accordingly the former relative error is significantly smaller than the later, from which we can see that the amount of training data has an important impact on the algorithm accuracy.

Under downlink EFR mode, the accuracy difference between training and testing of the algorithm is larger in MOS interval of (1, 2] because the amount of available data in the interval is smaller, leading to local instability of the algorithm.

For FR and HR modes, because of less data and poor MOS ergodicity, it is still not sufficient to support effective training of the algorithm.

In conclusion, when the amount of training data is adequate and the distribution of MOS value is evenly, the algorithm provides a high measuring accuracy.

4. Conclusion

This paper proposed a QoE measuring algorithm of voice service for GSM network, taking radio link parameters which can be obtained from mobile network in real time as inputs. Multiple regression and principle component analysis are combined in the modeling approach of QoE assessment. The method is especially convenient to be integrated into signaling monitoring platform of wireless networks. Both the algorithm's training and testing procedures use data collected from commercial GSM networks, and the result has shown that with adequate valid data, the algorithm will achieve high accuracy. Furthermore, the proposed QoE prediction method based on GSM network can also be extended to other wireless networks such as Universal Mobile Telecommunications System (UMTS) and Long-term Evolution (LTE).

REFERENCES

[1] ITU-T P.10/G.100, "Vocabulary and Effects of Transmission Parameters on Customer Opinion of Transmission Quality," 2008.

[2] ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality," 1996.

[3] ITU-T Recommendation 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," 2001.

[4] Huawei Technologies Co., Ltd. "The Methods and Devices for the Estimation of Speech Quality," China Patent No. 200710172408.7, 2009.

[5] Ericsson Telefon AB-LM, "Speech Quality Measurement in Mobile Telecommunication Networks Based on Radio Link Parameters," US Patent No. 19970861563, 2000.

[6] Y. J. ZUO, "Perception of Voice, Win by Method—Solutions for Voice Quality Assessment in TD-SCDMA by HUAWEI: VQI," *Mobile Communications*, Vol. 34, No. 3, 2010, pp. 30-31.

[7] ITU-T Recommendation P.862.3, "Application Guide for Objective Quality Measurement Based on Recommendations P.862, P.862.1 and P.862.2," 2007.

[8] M. Kantardzic (translated by S.Q. Shan, Y. Chen and Y. Cheng), "Data Mining," Tsinghua University Press, Beijing, 2003.

[9] ITU-T Recommendation P.862.1, "Mapping Function for Transforming P.862 Raw Result Scores to MOS-LQO," 2003.