QoS/QoE measurement system implemented on cellular phone for NGN

Satoshi UEMURA, Norihiro FUKUMOTO, Hideaki YAMADA, Hajime NAKAMURA
Communications Network Planning Laboratory
KDDI R&D Laboratories Inc.
2-1-15 Ohara Fujimino-shi Saitama 356-8502 Japan
{sa-uemura, fukumoto, hd-yamada, nakamura}@kddilabs.jp

***Abstract*—Services provided in Next Generation Network (NGN) must provide guaranteed quality end-to-end. This is quite a challenging issue since the fluctuation of network conditions is large in a Fixed Mobile Convergence (FMC) network. Therefore, a novel approach where a network node and a mobile terminal such as a cellular phone cooperate with each other to control the service quality is essential. In order to achieve such cooperation, the mobile terminal should become more intelligent so as to be able to measure the service quality and notify the result to the network node. Then, the network node implements some kind of service control function such as a resource or admission control function based on the notification from the mobile terminal. In this paper, the role for the mobile terminal in such cooperated system is focused on and we describe a QoS/QoE measurement system implemented on a cellular phone. By using our implemented system, we can measure the user's perceptual quality of VoIP as well as the network QoS metrics such as packet loss rate, delay and jitter in real time.**

Keywords: speech quality, QoE/QoS, cellular phone, VoIP, RTP, NGN

I. INTRODUCTION

The remarkable benefit of the Next Generation Network (NGN) for users is that it offers a variety of services such as voice, text and video with high quality by using an IP-based network. IP communications are widely expanding into a variety of infrastructures based on mobile networks as well as fixed networks. The services provided in such FMC (Fixed Mobile Convergence) networks should maintain their quality end-to-end. This is one of the most crucial issues for making NGN successful.

In order to maintain end-to-end service quality, the mobile terminal such as a cellular phone and the network node should cooperate with each other as shown in Figure 1. In Figure 1, the mobile terminal plays the role of measuring the service quality and notifying the result to the network node. On the other hand, the network node plays the role of implementing resource or admission control based on the service quality notified from the mobile terminal. To achieve such cooperation, the mobile terminal must become more intelligent so as to be able to measure the service quality based on the user's perceptual quality.

Now, a cellular phone is an indispensable communication device since many attractive services as well as a voice service are provided on a cellular phone. Therefore, the cellular phone is regarded as a representative communication device in the NGN-era. From such a viewpoint, in this paper, we describe a QoS/QoE measurement system implemented on cellular phone. As for the method of measuring the user's perceptual quality, we propose an objective speech quality estimation method that is applicable to a cellular phone with poor computational ability.

This paper is organized as follows. In section II, the configuration and features of the QoS/QoE measurement system are summarized. In section III, an objective speech quality assessment that is a part of the measurement system is described. The performance of the implemented measurement system is verified in section IV, and concluding remarks are given in section V.

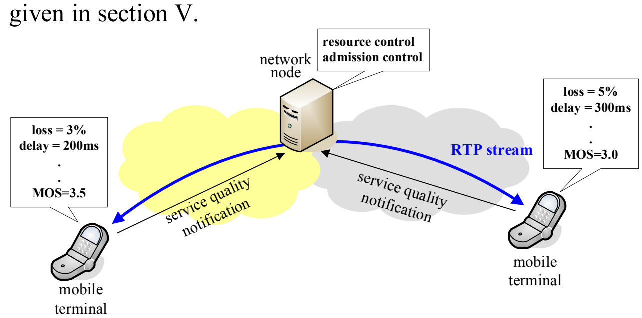


Figure 1: Cooperation between the network node and the mobile terminal

II. SYSTEM OVERVIEW

In this section, the configuration and features of the implemented system are described. Figure 2 shows an overview of our implemented system. This system mainly consists of SIP, RTP and QoS/QoE measurement components.

A. SIP

In this system, a VoIP session can be established based on Session Initiation Protocol (SIP) [1] and Session Description Protocol (SDP) [2]. The basic procedure from establishing to terminating the VoIP session is shown in Figure 3. Speech codec negotiation is also performed during this procedure.

1-4244-1457-1/08/$25.00 © IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.

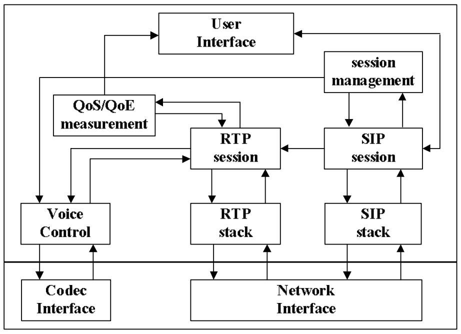


Figure 2: System overview

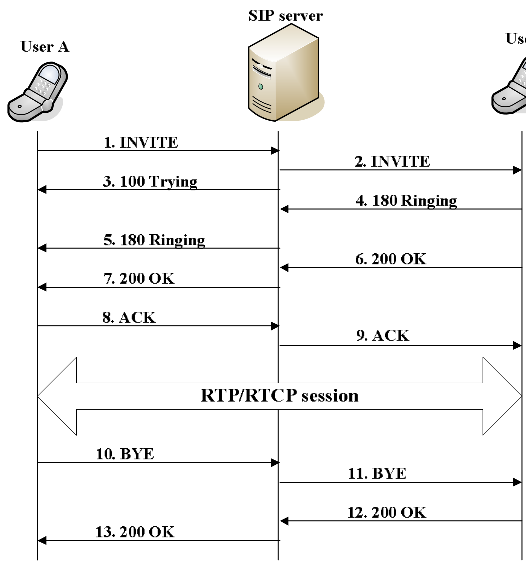


Figure 3: Basic procedure for establishing/terminating VoIP session

B. Speech Codec

In our system, two speech codecs, ITU-T Rec. G.711 [3] and ITU-T Rec. G.729 [4], can be used. Note that the bandwidth required by G.711 is 64 kbps, while the bandwidth required by G.729 is only 8 kbps. Based on the network condition, we can select which speech codec should be used.

C. RTP/RTCP

Once the VoIP session is established, the voice captured by the microphone is digitized and packetized. Then the digitized voice packets are sent by using Real-Time Transport Protocol (RTP) [5], which is widely used in VoIP systems.

RTP Control Protocol, RTCP, is implemented in this system. RTCP is an accompanying protocol to feedback the network conditions from the RTP receiver to the RTP sender. From the feedback report sent by the RTP receiver, the RTP sender can know the packet loss ratio, jitter, delay and so on that represent the network conditions.

The extended version of RTCP (RTCP XR) [6] is also implemented in this system. RTCP XR consists of seven report blocks so that the RTP sender can receive more detailed feedback from the RTP receiver. For instance, the packet format of the RTCP XR VoIP Metrics, which includes many metrics pertaining to the quality of the VoIP session, is shown in Figure 4.

BT=7	reserved	block length=8	
SSRC of source			
loss rate	discard rate	burst density	gap density
burst duration		gap duration	
round trip delay		end system delay	
signal level	noise level	RERL	Gmin
R factor	ext. R factor	MOS-LQ	MOS-CQ
RX config	reserved	JB nomial	
JB maximum		JB abs max	

Figure 4: RTCP XR VoIP Metrics

D. QoS/QoE measurement component

A QoS/QoE measurement component calculates the QoS metrics and the QoE metrics by using RTCP and RTCP XR feedback reports. In this paper, the term “QoS metrics” denotes metrics regarding network conditions, while “QoE metrics” denotes metrics regarding the user’s perception. As QoS metrics, packet loss rate, delay, jitter and so on are calculated in this component. As QoE metrics, the R factor [7], MOS-LQ (MOS Listening Quality) and MOS-CQ (MOS Conversational Quality) and so on are calculated. The R factor can be calculated by using about 20 QoS metrics such as packet loss rate and delay, and is expressed as a scalar in the range of 0 to 100. MOS-LQ and MOS-CQ are expressed as a scalar in the range of 1 to 5. In the case where the speech quality remains good, MOS-LQ and MOS-CQ show a score near 5. In our QoS/QoE measurement component, MOS-LQ can be calculated by using our proposed method described in section III, while MOS-CQ can be calculated by well-known method using the R factor as follows [7].

$$\mathrm{MOS-CQ} = \begin{cases} 1 & , R < 0 \\ 1 + 0.035R + R(R-60)(100-R)7\cdot 10^{-6} & , 0 < R < 100 \\ 4.5 & , R > 100 \end{cases} \quad (1)$$

III. ESTIMATION OF END USER’S QOE OF VOIP

The most reliable metric of the service quality at the end point is the user’s perceptual quality. Mean Opinion Score (MOS) defined by ITU-T Rec. P.800 [8] is widely used as the user’s perceptual quality of speech. However, in terms of in-service speech quality, MOS cannot be measured, because MOS is the statistical average of a large number of user’s subjective opinion scores which are integers from 1 to 5. Therefore several objective speech quality assessments have been developed to make a good estimation of MOS.

Objective speech quality assessment can be classified into two types; one is intrusive and the other is non-intrusive. The intrusive type requires the reference speech to estimate the MOS, while the non-intrusive type does not require this. Since generally the reference speech cannot be obtained in an in-service VoIP session, the non-intrusive type is desirable to be implemented on a cellular phone. In addition, the processing performance of a cellular phone is not sufficiently high to

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.

perform signal processing that requires high computational cost. Therefore here, an objective MOS-LQ estimation method based on ITU-T Rec. P.564 [9] with low computational cost is proposed.

The proposed estimation method can calculate MOS-LQ simply using a set of QoS metrics that can be observed in an IP-layer. Since most of the QoS metrics are included in a RTCP or RTCP XR report, redundant calculations for estimating the speech quality such as signal processing for frequency analysis are not necessary in the proposed method.

An overview of the proposed MOS-LQ estimation method is shown in Figure 5. Note that though a packet generally consists of several frames, in this paper we assume a packet length is 20msec so that a packet consists of one 20msec frame. First of all, the packet loss is detected according to the "sequence number" field of the RTP header. Next, the payload of the lost packet is discriminated in terms of whether the packet includes a voice part or not. If the packet includes a voice part, the end user perceives that the voice quality is seriously degraded. On the other hand, if the packet includes only a silent part, the user rarely perceives the quality degradation. In the proposed method, the normalized power of the most recent arrival packet is used to determine whether the lost packet includes the voiced part or not. From such discrimination, each packet loss is weighted based on its significance as regards the user's perception.

Next, the metrics that are likely to affect the speech quality are calculated. Here we choose the six metrics that are likely to affect speech quality as shown in Table 1; the packet loss rate is obviously important factor for estimating the speech quality and its pattern, which is random or bursty, is also important [10]. Since we assume that MOS-LQ can be represented by a linear combination of the six metrics, we applied multiple linear regression analysis to determine the relation between these six metrics and MOS-LQ; we retained the partial regression coefficients for each metric in advance. Finally, the estimated MOS-LQ can be obtained by using the six metrics and the partial regression coefficients as follows.

$$\begin{aligned}\mathrm{MOS-LQ} = {} & \alpha_1 \cdot \mathrm{loss\_rate} + \alpha_2 \cdot \mathrm{burst\_duration} \\ & + \alpha_3 \cdot \mathrm{gap\_duration} + \alpha_4 \cdot \mathrm{burst\_density} \\ & + \alpha_5 \cdot \mathrm{gap\_density} + \alpha_6 \cdot \mathrm{R\_factor} + \alpha_7\end{aligned} \quad (2)$$

where $\alpha_i \ (i = 1, 2, \ldots, 7)$ is a partial regression coefficient for each metric which can be determined by multiple linear regression analysis.

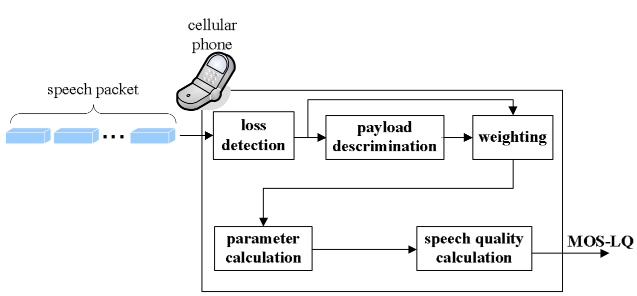


Figure 5: Overview of proposed method

Table 1: Metrics used in proposed method

Metric	Summary
loss rate	Packet loss rate
burst duration	Mean duration of burst periods
gap duration	Mean duration of gap periods
burst density	Fraction of RTP data packets within burst periods
gap density	Fraction of RTP data packets within gap periods
R factor	R factor defined in [7]

IV. PERFORMANCE EVALUATION OF IMPLEMENTED SYSTEM

In order to verify this system, performance evaluation tests were conducted. The test conditions were as follows.

A. Test conditions

A performance evaluation test was conducted as shown in Figure 6. The sender sends speech data to the receiver through an access point for wireless LAN (IEEE802.11b) and a network simulator. The network simulator can simulate the packet loss, delay and jitter of a real network. It is widely known that packet loss distribution in IP networks can be modeled by using a Markov process [11]. In this test, the packet loss was generated by using the Gilbert-Elliott model [12].

As shown in Figure 7, the Gilbert-Elliott model is a two state Markov model that consists of "Good" and "Bad" states. In the "Good" state, packet loss occurs according to a probability P_G, which is almost zero, while in the "Bad" state, packet loss occurs according to a probability P_B. In this test, we set $P_G = 0$ and $P_B = 0.5$. Here, we designate the parameters that represents the burstiness of packet loss and packet loss rate as $\gamma (0 \le \gamma \le 1)$, f, respectively. Then, the state transition probabilities p and q are given as follows:

$$p = (1-\gamma)\left(1 - \frac{P_B - f}{P_B - P_G}\right) \quad (3)$$

$$q = 1 - p - \gamma \quad (4)$$

The parameters that should be set for the model, γ and f, were set as shown in Table 2. When $\gamma = 0.2$, packet loss distribution will be random, while when $\gamma = 0.8$, packet loss distribution will be bursty.

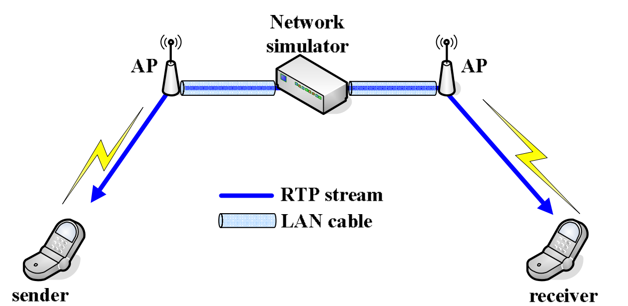


Figure 6: Configuration of performance evaluation test

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.

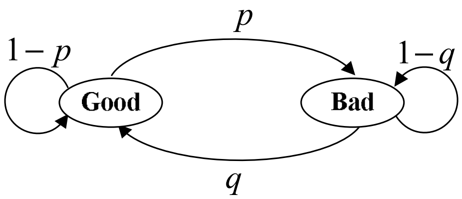


Figure 7: Gilbert-Elliott model

Table 2: Parameter configuration of Gilbert-Elliott model

Parameter	Value
Burst index γ	0.2, 0.8
Packet loss rate f	0～10% (every 1%)

In this test, 40 Japanese sentences, each of them 8 seconds long, were used as original speech data. First, these speech data were sent by the sender's cellular phone to the receiver's cellular phone as shown in Figure 6. Then, the speech data were degraded due to packet loss generated by the network simulator. Note that the packet loss is generated by using the Gilbert-Elliott model as described above. Next, the receiver recorded the degraded speech data in the cellular phone. Then, we evaluated the degraded speech data objectively by using our proposed speech quality assessment described in section III and ITU-T Rec. P.862.1 [13]. Note that MOS-LQ calculated by using ITU-T Rec. P.862.1, denoted by MOS-$LQ_{p862.1}$, has a strong correlation with subjective MOS. Therefore if MOS-LQ calculated by using our proposed method, denoted by MOS-LQ_{prop}, has a strong correlation with MOS-$LQ_{p862.1}$, it follows that MOS-LQ_{prop} also has a strong correlation with subjective MOS.

Since the Gilbert-Elliott model simulates the packet loss event according to the state transition probability, the generated packet loss distribution is different in each trial, even if we use the same pair of γ and f. Therefore we conducted the procedure to obtain degraded speech data as described above three times, and consequently three data sets A, B, C were obtained. Note that each data set includes 880 degraded speech data, since two patterns of γ and eleven patterns of f for 40 sentences generate 880 degraded speech data.

B. *Peformance evaluation result*

Figure 8 and Table 3 show the relation between MOS-LQ_{prop} and MOS-$LQ_{p.862.1}$. It can be seen from Figure 8 that the majority of points lie on the dashed diagonal line. In this case, the correlation coefficient between MOS-LQ_{prop} and MOS-$LQ_{p.862.1}$ is higher than 0.87. Furthermore, we can see from Table 3 that MOS-LQ_{prop} has a strong correlation with MOS-$LQ_{p.862.1}$ regardless of the codec type and data set.

Figure 9 shows the estimation error distribution of the proposed method. It can be seen from Figure 9 that almost 84% of the test data are within the error range ± 0.25 and almost 97% are within the error range ± 0.5, respectively. Consequently, the speech quality at the endpoint can be measured with high reliability by using our proposed method.

C. *GUI of measurement result*

The QoS/QoE measurement results on the cellular phone are shown in Figure 10. In Figure 10 (a), the number of sent/received packets are shown in field (A). A speech quality indicator is shown in field (B) of Figure 10 (a). This indicator shows a green, yellow and red lamp according to the speech quality. While the speech quality remains good, this indicator shows a green lamp, but if the speech quality is degraded, it turns red from green through yellow. Figure 10 (b) shows the indicator in the case of the speech quality becoming worse. Round Trip Time, R factor, MOS-LQ and MOS-CQ are displayed in field (C) of Figure 10 (a). In this field, the terms "own" and "rem" mean own cellular phone and remote cellular phone, respectively. Figure 10 (c) shows a detailed quality information window. This window includes the feedback report of RTCP XR.

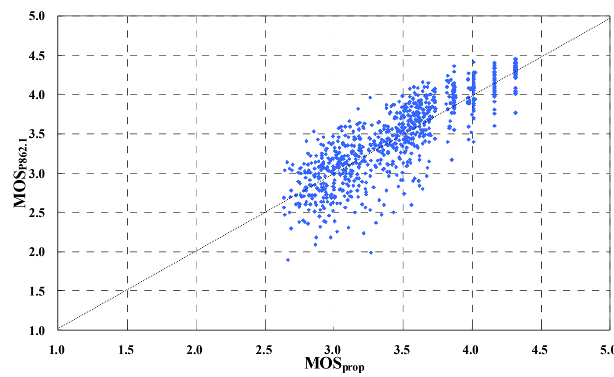


Figure 8: Estimation accuracy

Table 3: Correlation coefficient between $MOS_{p862.1}$ and MOS_{prop}

codec	data set A	data set B	data set C
G.711	0.87	0.86	0.87
G.729	0.87	0.86	0.86

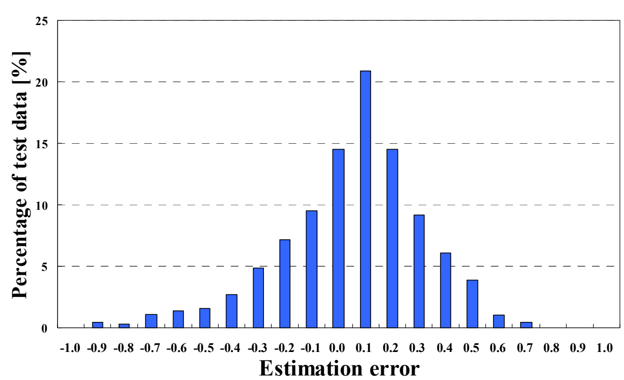
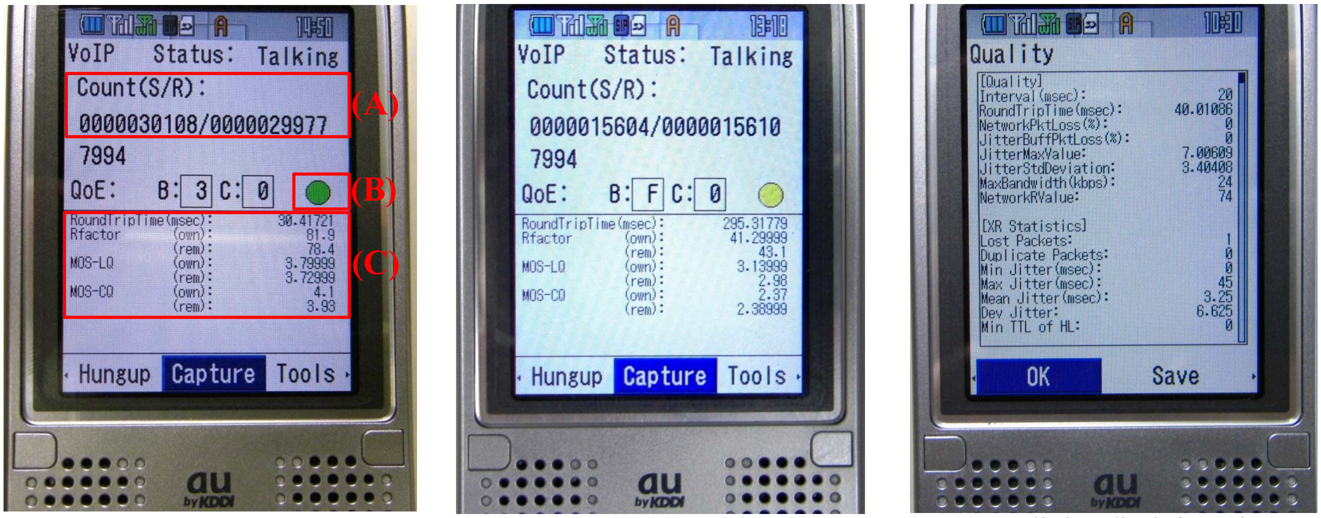


Figure 9: Estimation error distribution



(a) A QoS/QoE measurement result in good speech quality

(b) A QoS/QoE measurement result in bad speech quality

(c) Detailed quality information window

Figure 10: QoS/QoE measurement result on screen

V. CONCLUSIONS

In this paper, we described a QoS/QoE measurement system implemented on a cellular phone. By implementing the QoS/QoE measurement system, we can measure a user's perceptual quality as well as the network conditions in real-time. As for the method of measuring the user's perceptual quality, we proposed an objective speech quality estimation method that is applicable to a cellular phone with poor computational ability. In spite of the simple algorithm, experimental results show that our proposed method can estimate the speech quality with high accuracy regardless of the codec type.

The service quality measured by a cellular phone can be used as a trigger to activate the resource and admission control function of the network node. Two scenarios can be envisaged. One scenario is that the cellular phone sends a service control request to the network node when the service quality falls below a certain level. Then, the network node implements resource or admission control based on the request from cellular phone. The other scenario is that the network node monitors the service quality notified from the cellular phone, and when the service quality falls below a certain level, the network node implements resource or admission control.

In order to develop such cooperation between the network node and cellular phone, we will study the service control scheme in the network node in further detail in future.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Dr. M. Suzuki and Dr. S. Akiba of KDDI R&D Laboratories, Inc. for their support of our work. This work was partly supported by the National Institute of Information and Communications Technology (NICT).

REFERENCES

[1] J. Rosenberg, et al, "SIP Session Initiation Protocol," IETF RFC 3261, June. 2002.

[2] M. Handley and V. Jacobson, "SDP: Session Description Protocol," IETF RFC 2327, April 1998.

[3] ITU-T Recommendation G.711, "Pulse code modulation (PCM) of voice frequencies," International Telecommunication Union, November. 1988.

[4] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using conjugate structure algebraic codec excited linear prediction," International Telecommunication Union, March. 1996.

[5] H. Schulzrinne, S. Casner, R. Frederic, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 3550, July 2003.

[6] T. Friedman, R. Caceres, and A. Clark, "RTP Control Protocol Extended Reports (RTCP XR)," IETF RFC 3611, Nov. 2003.

[7] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning," International Telecommunication Union, May 2003.

[8] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality,"International Telecommunication Union, Aug. 1996.

[9] ITU-T Recommendation P.564, "Conformance testing for narrowband voice over IP transmission quality assessment models," International Telecommunication Union, July 2006.

[10] A. Clark, "Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality," in proc. IPtel 2001, pp.123-127, April 2001.

[11] ITU-T Recommendation G.1020, "Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks," International Telecommunication Union, July 2006.

[12] ITU-T Recommendation G.191, "Software tools for speech and audio coding standardization," International Telecommunication Union, Nov. 2000.

[13] ITU-T Recommendation P.862.1, "Mapping function for transforming P.862 raw result scores to MOS-LQO," International Telecommunication Union, Nov. 2003.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.