VoIP 통화 품질 평가를 위한 개선된 E-모델

Advanced E-Model for VoIP Call Quality Assessment

최승권*, 송종명*, 이병록**, 황병선***, 조용환*

충북대학교*, (주)유비컴테크놀로지**, (주)제넥스웨이브***

Seung-Kwon Choi(hamsterc@hanmail.net)*, Jong-Myeong Song(sadness71@hanmail.net)*, Byeong-Rok Lee(bblsh1689@hanmail.net)**, Byeong-Seon Hwang(rhayoon@hanmail.net)***, Young-Hwan Cho(yhcho@chungbuk.ac.kr)*

요약

본 논문에서는 기존의 문제점들을 극복하기 위한 개선된 E-Model을 제안하였다. 새로운 모델은 버스트 패킷손실과 최신효과를 적용하여 VoIP 품질 측정의 정확도를 높일 수 있다. 개선된 E-Model은 길버트 모델에 의해 생성된 버스트 패킷 손실에 따라 NR(Network R) 값과 UR(User R) 값을 측정한다. 기존 모델들인 MOS, PESQ, E-Model과 비교하여 시뮬레이션을 수행하였으며, 실험 결과 개선된 E-Model이 기존 모델들보다 정확하고 신뢰성 있음이 증명되었다.

■ 중심어 : | E-Model | MOS | PESQ | VoIP | 통화품질 |

Abstract

In this paper, an advanced E-Model was proposed in order to overcome disadvantages of conventional method. A new model makes the accurate VoIP call quality assessment possible by applying the burst packet loss and recency effect. In order to assess the performance of this advanced E-Model, we gained the estimated MOS value from NR(Network R) value and UR(User R) value resulted from the burst packet loss values by Gilbert Model. Through simulations and comparisons with conventional models such as MOS, PESQ, and E-Model, we reach a conclusion that advanced E-Model is more accurate and reliable method than conventional models.

■ keyword : | E-Model | MOS | PESQ | VoIP | Call Quality |

I. Introduction

As the number of Internet users has increased extremely, a various of studies and competitions have been accelerated for the development of Internet. VoIP service(Voice over IP) that has recently come into prominence also has become an

* 이 논문은 2005년도 충북대학교 산학협력단 부설 유비쿼터스바이오정보기술연구센터 지원에 의하여 연구되었음(This work was supported the Research Institute of Ubiquitous Bio-information Technologe in 2005).

접수번호 : #050707-001

접수일자 : 2005년 07월 07일

심사완료일 : 2005년 08월 13일

교신저자 : 조용환, e-mail : yhcho@chungbuk.ac.kr

interesting Internet application. And as VoIP services are popularized, customers using VoIP service are demanding the call quality that is similar to the existing PSTN service.

There are various elements that influence the call quality of VoIP service such as delay, packet loss and jitter in IP network including codec delay, echo in the VoIP terminal. Therefore, VoIP service enterprisers are assigned to reduce such losses by analyzing various services quality elements in order to provide the high quality of VoIP service.

The first method that assess the call quality is MOS(Mean Opinion Score). MOS is the method that assess call quality from mean opinion score of call quality by using many assessor, and this method is rather an intuitional assessment but is not appropriated for accurate quality assessment because of high subjectivity. To overcome this subjectivity, some objective methods such as PSQM(Perceptual Speech Quality Measurement), PAMS(Perceptual Analysis Measurement System) and PESQ(Perceptual Evaluation of Speech Quality) were newly introduced. However, these models are assessing only the quality of perceptual voice data based on transmitted voice and thus, the result of assessment is varied with each voice sample. Also such methods are inappropriate to reflect the network environmental elements that have impact upon the voice quality. Therefore the E-Model was introduced.

II. VoIP service quality elements.

There are a lot of elements that affect VoIP service quality. For example, codec delay or echos in VoIP terminal, delay in IP networks, packet loss and jitter are prominent elements decreasing VoIP service quality[1].

VoIP service quality are generally classified to the connection quality and call quality.

1) Connection quality

- Time to set up.
- Time to connection between gatekeeper and gateway.
- Time to connect to the server to process certificate fees, duties and taxes etc.

2) Call quality

- End-to-end delay : codec delay + packet process and buffering + network transmission delay.
- Impacting elements for call quality : coding noise, echo, packet loss, lack of bandwidth.

III. Advanced E-Model for VoIP call quality assessment

3.1 E-Model

The complexity of modern networks requires that for transmission planning the many transmission parameters are not only considered individually but also that their combination effects are taken into account. This can be done by "expert, informed guessing," but a more systematic approach is desirable, such as by using a computational model. This computational model can be useful to transmission planners, to help ensure that users will be satisfied with end-to-end transmission performance.

The output from the model is a scalar quality rating value, R, which varies directly with the overall conversational quality. However, the output can also give nominal estimates of user reactions, for instance in the form of percentages finding the modelled connection "Good or Better" or "Poor or Worse".

The primary output of the model is a scalar rating of transmission quality. It must be emphasized that the primary output from the model is the "Rating Factor" R but this can be transformed to give estimates of customer opinion. Such estimates are only made for transmission planning purposes and not for actual customer opinion prediction. A major feature of this model is the use of transmission impairment factors that reflect the effects of modern signal processing devices.

The result of any calculation with the E-Model in a first step is a transmission rating factor R, which combines all transmission parameters relevant for the considered connection. This rating factor R is composed of

$$: \quad R = Ro - Is - Id - Ie\text{-}eff + A \qquad (1)$$

Ro represents in principle the basic signal-to-noise ratio, including noise sources such as circuit noise and room noise. The factor *Is* is a combination of all impairments which occur more or less simultaneously with the voice signal. Factor *Id* represents the impairments caused by delay. *Ie-eff*, effective equipment impairment factor, represents impairments caused by low bit rate codecs. It also includes impairment due to packet-losses of random distribution. The advantage factor A allows for compensation of impairment factors when there are other advantages of access to the user. The term *Ro* and the *Is* and *Id* values are subdivided into further specific impairment values.

3.2 Advanced E-Model

With respect to the application of the E-model to the planning of modern networks, however, there are still several open problems, which limit the model's usability. They result from the terminal and transmission characteristics of modern networks, which could not have been taken into account at the time the model was established. In order to guarantee that the achieved concept keeps track with the technological progress in both the transmission system and the terminal equipment areas, it is highly desirable to maintain and update the model. The validity range of the E-model should be extended so that it cannot only be applied to traditional networks, but packet-based transmission, wide-band systems, or non-handset terminal equipment.

VoIP service has time varying call quality in the following [Fig. 1]. Generally, VoIP service quality are classified to the connection quality and call quality. Connection quality includes time to set up, time to connection and call quality includes end-to-end delay, coding noise, echo, packet loss, lack of bandwidth and so on.

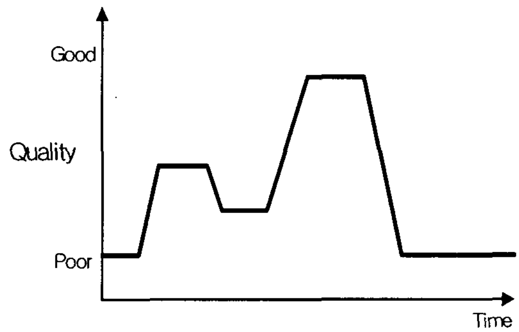

Fig. 1. Time varying call quality

But. conventional E-Model didn't consider the effects of time varying impairments such as burst packet loss and recency effect.

3.2.1 Burst packet loss

The network impairment that has greatest effect on voice quality is packet loss. Packet loss may occur due to buffer overflow within the network, deliberate discard as a result of some congestion control scheme or transmission errors.

The equipment impairment factor *Ie* in E-Model is generally used to represent the effects of Voice over IP equipment. Certain codecs have been characterized through subjective testing to give a profile of the variation of *Ie* with packet loss in the following [Fig. 2][2].

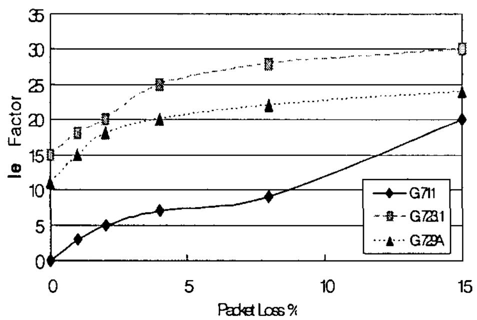

Fig. 2. Mapping packet loss to Ie

Several of the mechanisms that can lead to packet loss are of a transient nature and hence the resulting packet loss is bursty in nature as shown [Fig. 3] Low packet loss rates a burst distribution gave a higher subjective quality than a non-bursty distribution whereas for high packet loss rates the converse was true as shown [Fig. 4].

Bolot studied the distribution of packet loss in the Internet and concluded that this burst packet loss could be represented by a Markovian loss model such as the Gilbert or Elliott models[3].

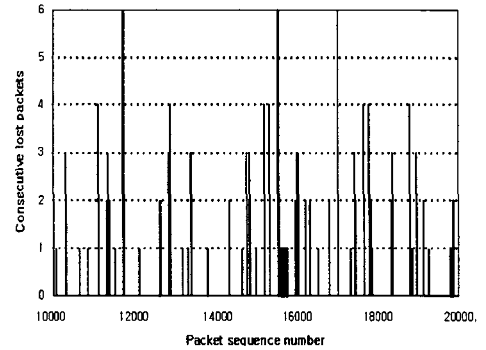

Fig. 3. Burst packet loss

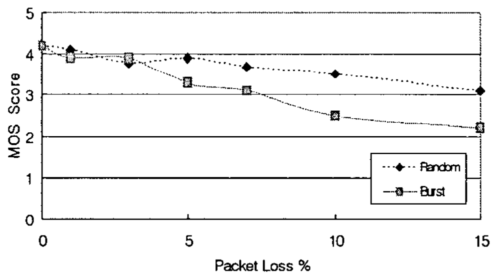

Fig. 4. Effects of random vs burst packet loss

In conclusion, packet loss concealment is effective for isolated lost packets but can't hide periods of high packet loss, i.e. 1% random packet loss may be ok but 1% burst packet loss may be problem. This case may be more critical problem in periods of high packet loss rate.

3.2.2 Recency effect in VoIP call quality measurement

If the rate of packet loss varies during a VoIP call then the perceived call quality will also vary. The term 'instantaneous quality' may be used to denote the measured or calculated quality due to packet loss or other impairments and the term 'perceived quality' may be used to denote the quality that the user would report at some instant in time.

Intuitively, if instantaneous quality changes from 'Good' to 'Bad' at some moment in time then the listener would not immediately notice the change. As time progresses the user would become progressively more annoyed or distracted by the impairment. This leads to the idea that the perceived quality changes more slowly than instantaneous quality as shown [Fig. 5].

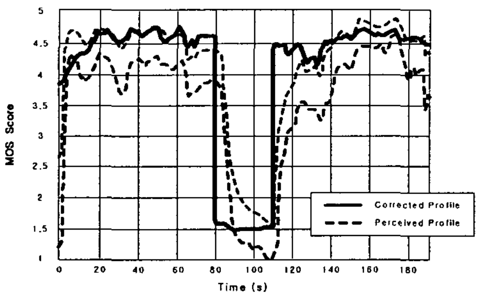

Fig. 5. Instantaneous and Perceived quality metrics(Source Barriac[4])

In tests reported by Barriac et al the packet loss rate during a 3 minute call was varied from 0 to 25%. In the above [Fig. 5], the packet loss was set to 25% for most of the call and reduced to 0% for a 30 second period mid-call. Listeners were asked to move a slider to indicate their assessment of quality during the call and then asked to rate the call at the end. This showed the recency effect with an approximately exponential curve with a time constant of 5 seconds for the good-to-bad transition and 15 seconds for the bad-to-good transition[4].

This effect reflects the way that a listener would remember call quality. When the noise was at the start of the call users reported a MOS score of 3.82 whereas when the noise was at the end of the call users reported a MOS score of 3.18, giving a change in MOS score of 0.64. This recency effect is believed to be due to the tendency for people to remember the most recent events or possibly due to auditory memory which typically decays over a 30 second interval.

This recency effect induce the difference between the real call quality and measured call quality as shown [Fig. 6][5].

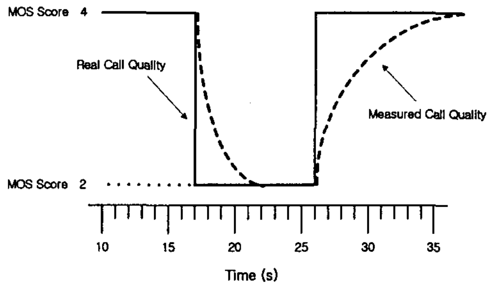

Fig. 6. Real call quality and Measured call quality

3.2.3 Advanced E-Model

This advanced E-Model is based on conventional E-Model and considering burst packet loss and recency effect. The block diagram of advanced E-Model is shown in [Fig. 7]. As the result of measurement, 'NR Factor(Network R factor)' and 'UR Factor(User R Factor)' are received.

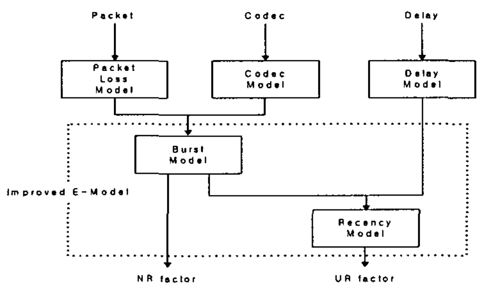

Fig. 7. Block diagram of advanced E-Model

The conventional E-Model can then be represented as :

$$R = Ro - Is - Id - Ie\text{-}eff + A \qquad (2)$$

In many VoIP network implementations the connection between codec and telephone handset may be transient. For example a user may be dialing through an existing local loop and being routed to a Gateway located at the central office. This means that some elements of the E-Model may not be measurable by equipment located within the network. Default values for many of the E-Model parameters can be assumed as shown in the past research, giving an effective value for *Ro* - *Is* of 93.2[6].

So generally E-Model can then be simply represented as :

$$R = 93.2 - Id - Ie \qquad (3)$$

Id represents the impairments caused by delay and Ie represents the impairment due to random packet losses.

1) delay impairment *Id*

The effects of delay are well known and easily modeled. Delays of less than 175mS have a small effect on conversational difficulty whereas delays over 175ms have a larger effect. A simply delay model is following code.

IF delay < 175 mS then
Id = 4.delay
Else
Id = 4 + (delay - 175) / 9

2) NR factor

The average value of *Ie* may be determined by taking the average of the perceived quality for the call. For each time interval $t(i)$, the instantaneous quality $I_{inst}(i)$ is determined by measuring the post-jitter buffer packet loss for the time interval and mapping the packet loss to an *Ie* value using the curves shown in [Fig. 2].

Over a series of *N* samples the average perceived quality is therefore

$$Ie(av) = sum(I_{perceived}(i)) / N \qquad (4)$$

In a 2-state Markov model represent the burst packet loss characteristics of the call. The two states represent the conditions of receiving or losing a packet within burst or silent states.

State 1 - Silent state
State 2 - Burst state

The equipment impairment value for the burst and silent state is determined using the curves shown in [Fig. 2] giving I_{eb} and I_{eg} respectively.

Let I_1 be the quality level at the change from burst state I_{eb} to silent state I_{eg} and let I_2 be the quality level at the change from I_{eg} to I_{eb}.

The perceived quality can be estimated from the instantaneous quality by assuming an exponential decay, modeling the effect described in chapter 3.2.1. A time constant of 5 seconds is assumed for a deterioration in quality and 15 seconds for an improvement in quality.

As considering above explanation, the NR factor is calculated in the following steps.

$$I_1 = I_{eb} - (I_{eg} - I_2)e^{-b/t1} \qquad (5)$$

where t_1 is typically 5

$$I_2 = I_{eg} + (I_1 - I_{eg})e^{-g/t2} \qquad (6)$$

where t_2 is typically 15

Combining these gives

$$I_2 = (I_{eg}\ (1-e-g/t_2) + I_{eb}(1-e-b/t_1)e-g/t_2) \\ /(1-e-b/t_1-g/t_2) \quad (7)$$

Integrating the expressions for I_1 and I_2 to give a time average gives

$$Ie(av) = (b\ I_{eb}+g\ I_{eg}-t_1(I_{eb}-I_2)(1-e-b/t_1) \\ +t_2(I_1-I_{eg})(1-e-g/t_2)/(b+g) \quad (8)$$

This may be used to determine an NR factor from the expression :

$$NR = 93.2 - Ie(av) \quad (9)$$

This NR factor does not yet include the effects of delay or recency however is useful when examining the effects of packet loss, jitter and codec type on transmission quality.

3) UR factor

The recency effect can be modeled by assuming that perceived quality decays exponentially with time constant t from the 'exit' value I_{exit} from a burst of noise or distortion towards the average Ie. The following model is proposed :

$$Ie(end\ of\ call) = Ie + (k(I_{exit} - Ie))e^{-y/t3} \quad (10)$$

It is assumed that the I_1 represents the exit value from the last significant burst of packet loss, y represents the time delay since the last burst, t_3 is a time constant of typically 30-60 seconds and k is a constant.

$$Ie(end\ of\ call) = Ie(av) + (k(I_1 - Ie(av)))e^{-y/t3} \quad (11)$$

The UR factor is determined from the expression below. This is intended to more closely approximate the user's perspective of quality and therefore does take into account both recency and delay.

$$UR = 93.2 - Ie(end\ of\ call) - Id \quad (12)$$

Ⅳ. Performance analysis of advanced E-Model

4.1 Experiment environment

In order to assess the performance of the proposed advanced E-Model, a 2-state Gilbert Model using Markov chain is used to represent the burst packet loss characteristics of the call

The four conditions represent the cases of receiving or losing a packet within burst or silent state.

Condition 1 : Silent state-receive packet
Condition 2 : Burst state-receive packet
Condition 3 : Burst state-lose packet
Condition 4 : Silent state-receive packet

A silent state is defined by the requirement that Smin successive packets must be received. This model includes a condition representing the loss of an isolated packet within a silence. The rationale for this is that packet loss concealment can mask the effects of isolated lost packets.

Burst packet loss event :

```
C5 = C5 + npr
if npr >= Smin then
        if lost = 1 then
                C14 = C14 + 1
        else
                C13 = C13 + 1
```

```
        lost = 1
        C11 = C11 + npr
else
        lost = lost + 1
        if lost > 8 then C5 = 0
        if npr = 0 then
C33 = C33 + 1
else
                C23 = C23 + 1
                C22 = C22 + npr
npr = 0
```

npr is an input parameter representing the number of packets received since the last lost packet event. The series of counters *C11* to *C14* are used to determine the corresponding Markov model transition probabilities (i.e *C11* is used to calculate *P11*). Counter *C5* is used to measure the delay since the last 'significant' burst of lost packets. Parameter S_{min} is the minimum silent size.

A packet loss event driven model is used to count a minimum number of key transition events. It is assumed that voice activity detection is being used and hence that packet loss reports relate to packets containing speech energy. When the call is completed then remaining transition counts can be derived and then the counts normalized to give probabilities. This model holds considerable information and can be used to determine average silent and burst size and density, successive lost packet distribution etc.

In order to assess the performance of the proposed advanced E-Model, we gained the result value NR and UR factor from corrupted audio files using a burst error process by using proposed model.

NR factor is compared to MOS, PESQ and E-Model using each algorithm. And UR factor is compared to MOS and E-Model using each algorithm. These algorithms are offered by ITU-T.

4.2 Experiment results

Some initial subjective comparison was made to validate the advanced E-Model. An audio file was corrupted using a burst error process which comprised the loss and state transition probabilities being selected randomly. A 10mS packet size was used and packet loss concealment applied.

The fifty sets of 5 test audio files were created, and a group of six listeners used to rank the files from 1(best) to 5(worst). The ranking was compared with that predicted by the algorithm described above.

Table 1. Comparison of UR factor and User Ranking-File set 1

File	Mean opinion score	UR factor rank
1-1	1.0	1
1-2	3.0	2
1-3	3.5	3
1-4	2.5	4
1-5	5.0	5

Table 2. Comparison of UR factor and User Ranking-File set 2

File	Mean opinion score	UR factor rank
2-1	1.2	1
2-2	1.8	2
2-3	3.2	3
2-4	3.8	4
2-5	5.0	5

The results showed reasonable correlation with MOS however there were some exceptions, for example file 1-4. The locations of packet loss events for this file was reviewed and it became apparent that some loss bursts occurred either during silence periods or during periods when the sound produced by the speaker was not changing significantly.

4.2.1 Comparison with NR factor and result values of other model

The E-model provides a statistical estimation of quality measures. An estimated Mean Opinion Score MOSCQE for the conversational situation in the scale 1-5 can be obtained from the R factor. This R factor can be inverted in the range 6.5≤R ≤100 to calculate R from MOSCQE.

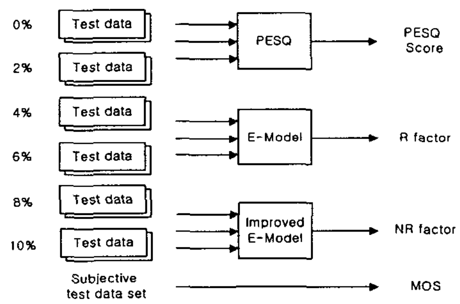

Fig. 8. Accepting result values for each model

The NR factors from the results of file set above were inverted to estimated MOS values and compared to result values of MOS, PESQ and E-Model in the period of packet loss rate 0-10%. [Fig. 9]. shows the result that compared with estimated MOS score of each model.

It shows that the NR value from advanced E-Model using burst packet loss model is similar to PESQ considering combinations of factors such as filtering, variable delay, coding distortions and channel errors. The result shows that NR value is more accurate than PESQ value.

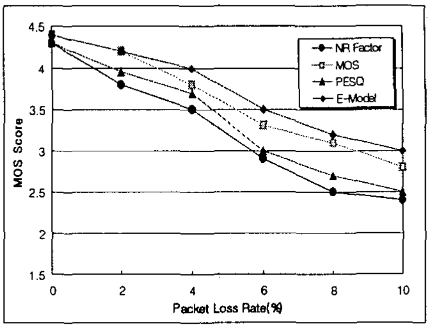

Fig. 9. Comparison with NR factor and other model values

4.2.2 Comparison with UR factor and result values of other model

The UR factors from the results of file set above were inverted to estimated MOS values and compared to result values of MOS and E-Model in the locations of packet loss event. The locations of packet loss event are shown in [Fig. 10].

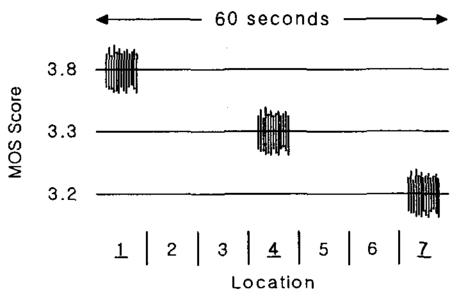

Fig. 10. The locations of packet loss event and recency effect

[Fig. 11] shows the result that compared with estimated MOS score of each model in the locations of packet loss event.

The result shows that the UR factor from advanced E-Model is reflecting time varying impairment such as recency effect. And it also shows that advanced E-Model outperform MOS and E-Model in predicting the relative subjective ranking of impaired audio files. The reason for this

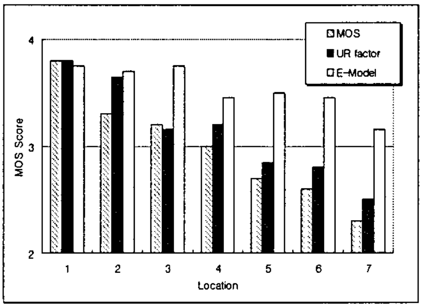

Fig. 11. Comparison with UR factor and other model values

is that the files were impaired using time varying impairments - MOS and E-Model assume that quality is relatively constant during a call whereas advanced E-Model is able to model the effects of time varying impairments such as recency effect.

V. Conclusion

This thesis states the problems of the conventional VoIP service assessment and proposes an advanced E-Model. This new model advanced two elements that was not considered in the conventional E-Model. First, the advanced E-Model reflected burst packet loss value to overcome the limit of random packet loss. In addition, time varying impairment values that ignored in the conventional E-Model can be calculated by the adoption of recency effect.

Through first simulation by packet loss model and comparison with conventional models, we knew that the NR value from advanced E-Model is much more accurate than MOS and E-Model. Also, we knew the NR value is similar or more accurate than PESQ considering combinations of factors such as filtering, variable delay, coding distortions and channel errors.

In addition, we knew that the UR factor from advanced E-Model is reflecting time varying impairment such as recency effect, and advanced E-Model outperform MOS and E-Model in predicting the relative subjective ranking of impaired audio files through second simulation and comparison.

Finally, we reached a conclusion that advanced E-Model is more accurate and reliable method than conventional models such as MOS, PESQ and E-Model.

In the future, VoIP service will break the time and space boundaries and will present in ubiquitous environment where wire and wireless are combined. Certainly, VoIP service used in ubiquitous environment will obviously be a critical issue in telecommunication industry. Therefore, it is recommended that a various of studies on VoIP service quality assessment should be continued in ubiquitous environment for the activation of VoIP service.

참 고 문 헌

[1] I. Lee, "The Standardization Tendency of Voice Service Quality," IT Standard Weekly, TTA, No.40, 2003.

[2] J. Seo and S. Woo, "A Study on the Application of an Amr Speech Codec for VoIP," IEEE, 2001.

[3] J. Bolot, "Adaptive FEC based Error Control for Interactive Audio in the Internet," Infocom, 1999.

[4] ITU SG 12 Contribution D.139, "The Relationship between Instantaneous and Overall Subjective Speech Quality for Time-varing Quality Speech Sequences," ITU, 2000.

[5] G. A. Carlesimo, "Recency Effect in Anterograde Amnesia," NeuroPsychologia, Pergamon Press, 2002.
[6] ITU-T Recommendation G.107, "The E-Model, a Computational Medel for Use in Transmission Planning," ITU-T, 2003.
[7] B. Douskalis, IP Telephony, PrenticeHall, Inc. 2000.

저 자 소 개

최 승 권(Seung-Kwon Choi) 정회원

- 2001년 8월 : 충북대학교 컴퓨터공학과 대학원 졸업(공학박사)
- 현재 : 충북대학교 초빙교수

<관심분야> : 멀티미디어 통신, 멀티미디어 컨텐츠 유통, 유비쿼터스 네트워킹

송 종 명(Jong-Myeong Song) 준회원

- 2005년 2월 : 충북대학교 컴퓨터공학과 대학원 졸업(공학석사)

이 병 록(Byeong-Rok Lee) 종신회원

- 현재 : 충북대학교 컴퓨터공학과 대학원 박사과정
- 현재 : (주)유비컴테크놀러지 대표이사

<관심분야> : 원격의료, e-learning contents

황 병 선(Byeong-Seon Hwang) 정회원

- 현재 : 충북대학교 컴퓨터공학과 대학원 박사과정
- 현재 : (주)제넥스웨이브 대표이사

<관심분야> : 멀티미디어

조 용 환(Young-Hwan Cho) 종신회원

- 1989년 2월 : 고려대학교 대학원(이학박사)
- 1982년 3월~현재 : 충북대학교 전기전자컴퓨터공학부 교수
- 현재 : (사)한국콘텐츠학회장

<관심분야> : 유비쿼터스 컴퓨팅, 멀티미디어 통신, 정보통신 정책