# PERCEIVED SPEECH QUALITY DRIVEN RETRANSMISSION MECHANISM FOR WIRELESS VoIP

Z Li, L Sun, Z Qiao and E Ifeachor
Department of Communication and Electronic Engineering
University of Plymouth, Plymouth, U.K.

***Abstract*—**Effective link Layer retransmission mechanisms in wireless networks are important as they can reduce packet loss due to bit errors. For wireless voice over IP (VoIP) , a key question that needs to be addressed in order to provide the best possible perceived speech quality is how to utilize retransmission schemes to recover corrupted packets whilst avoiding excessive retransmission delays. The contributions of this paper are two fold. First, we use an objective measure of perceived conversational speech quality (MOSc) as a metric to evaluate the performance of three current retransmission schemes (i.e. No Retransmission, Speech Property-Based Retransmission and Full Retransmission), while considering the impact of retransmission jitters. Our findings indicate that the performance of the retransmission mechanisms is a function of both wireless link quality and delay introduced in the wireline network. Second, we propose a new perceived speech quality driven retransmission mechanism which may be used to achieve optimum perceived speech quality for wireless VoIP (in terms of the objective mean opinion score) by switching to the most suitable retransmission schemes under different communication conditions.

## I.INTRODUCTION

Quality of Service (QoS) support for voice over IP (VoIP) in wireless/mobile networks is an important issue for technical and commercial reasons. However, speech quality for VoIP suffers from high packet loss rates and other impairments in the wireless link. Retransmission mechanisms, such as automatic repeat request (ARQ), have been incorporated in wireless and cellular networks to retransmit lost packets to improve performance in data transmission over wireless. In wireless networks such as 802.11b [1], the retransmission mechanism is a simple Stop & Wait algorithm and is implemented at the Media Access (MAC) layer, in which each transmitted packet must be acknowledged before the next packet can be sent. If in a certain timeout period an acknowledgement is not received by the sender of a frame, the sender will retransmit the frame until a maximal retransmission limit is reached. When the wireless link quality is poor, retransmission of MAC frames can effectively recover corrupted packets that contain bit errors.

However, excessive delays may be introduced by retransmission schemes that have significant adverse effects on real-time applications such as VoIP, which are sensitive to delay. A simplex retransmission scheme always negatively affects perceived speech quality in VoIP. There exists a tradeoff between packet loss and delay in a variety of retransmission schemes. Improved retransmission mechanisms such as Hybrid loss recovery scheme [2] and Speech Property-Based ARQ (SPB-ARQ) [3] have been proposed to reduce speech distortions by protecting packets that are perceptually more relevant. However, these schemes are only limited to listening-only quality assessment of the effect of the retransmission schemes on speech quality and do not consider the impact of delay which is important for conversation and interactivity. Further, these schemes do not consider the impact of retransmission jitters. Since adaptive jitter buffers would discard inappropriately retransmitted packets, the character of retransmission jitters introduced by different retransmission schemes should be considered.

The primary aim of the study reported in the paper is to investigate new retransmission mechanisms to improve speech quality for wireless VoIP. The contributions of the paper are twofold. First, we propose the use of a perceived conversational speech quality assessment method [4] to evaluate the performance of current retransmission mechanisms (No retransmission, Full retransmission, SPB retransmission) instead of listening-only method or individual network parameters (e.g. packet loss and delay). Second, we present a new retransmission policy, which can adapt to the most suitable retransmission mechanism, depending on the wireless link quality and network delay conditions. The ultimate aim of this perceived speech quality driven policy is to achieve optimum speech quality (in terms of the conversational Mean Opinion Score MOSc) in the face of network impairment factors and wireless channel situations, while considering the coupling effect of retransmission jitters and adaptive jitter buffers.

The paper is organized as follows, In Section II we describe the basic issues and methodology, including retransmission mechanisms, conversational speech quality evaluation and adaptive jitter buffers. Section III describes our simulation system. Results of simulations and the proposed perceived speech driven retransmission scheme is presented in Section IV. Section V concludes this paper.

## II.BASIC ISSUES AND METHODOLOGY

### A. Speech Property-based Retransmission Mechanisms

Speech Property-Based QoS control schemes are based on the fact that some voice frames are perceptually more important than others when encoded speech is transferred through packet networks. Recent experimental results show [5], that in some popular codecs used in wireless applications (e.g. AMR) the position of a frame loss has a significant influence on the perceived speech quality. In such codecs, frame loss concealment techniques are used to interpolate the parameters for the loss frames from the parameters of the previous frames. Lost voice frames at the beginning of a talkspurt will be concealed using the decoding information of previous unvoiced frames. However, because voiced sounds always have a higher energy than unvoiced sounds, concealment of these frames with unvoiced frames that have lower energy will cause a serious degradation in speech quality. Moreover, at the unvoiced/voiced transition stage, it is difficult for the decoder to correctly conceal the loss of voiced frames using the filter coefficients and the excitation for an unvoiced sound, especially when burst loss occurs or the frame size grows.

To maximise the perceptual quality at the receiving end, perceptually important voice packets may be protected by giving them a high priory with the unimportant packets handled as '*best-effort*'. For SPB retransmission, a retransmission scheme that protects only the perceptual important speech frames, is presented in [2][3]. Experimental results reported in [2] show that SPB retransmission could provides a better speech quality (assessed by EMBSD) than No retransmission scheme, which do not retransmit any packet. In [3], SPB retransmission was shown to be more efficient in reducing retransmission delays than Full retransmission, which retransmits every unacknowledged (unACKed) packet.

**B. MEASURING CONVERSATIONAL SPEECH QUALITY**

In previous studies [2][3], the assessment of retransmission schemes was performed using the EMBSD algorithm, which only considers the distortion caused by packet loss. However, in practice both packet loss and delay are crucial in voice conversation and long retransmission delays (e.g. due to long network delay) would seriously impact speech quality . The E-model [6] is introduced by ITU as a non-intrusive quality assessment method to obtain a measure of voice quality. Unfortunately, the E-model is only applicable to a limited number of codecs which at present does not include the AMR codec. In our simulation, we employed a technique that combines the PESQ and the E-model to evaluate the performance of different retransmission schemes. In the combined approach , the ITU PESQ is firstly used to quantify the impact of packet loss on speech quality. The result of this is then converted to the equipment impairment Ie. The average end-to-end delay effect, Id, is then calculated. The E-model is then used to obtain a measure of the speech quality, MOSc, based on Ie and Id (see Figure 1). Details of the implementation of the combined method are given in [4]

**C. Adaptive jitter buffer and Retransmission Jitters**

In VoIP applications, jitters are compensated for in the receiver by a jitter buffer. The size of a jitter buffer can be fixed or adjustable. Fixed jitter buffers cannot adapt readily to changes in network delays and as a result are not practical in real VoIP applications. In our study, we investigated *fast-exp*, one of the classical adaptive jitter buffer algorithms proposed in [7]. By using a smaller weighting factor as delays increase, the *fast-exp* algorithm can quickly adapt to the increases while avoiding discarding of too many packets. It estimates the current mean network delay (denoted as $\hat{d}_i$ ) and current variance of network delay (denoted as $\hat{v}_i$ ) when a packet arrives. The mean delay estimation equation is given by:

$$\begin{cases} \beta \hat{d}_{i-1} + (1-\beta) n_i : n_i > \hat{d}_{i-1} \\ a\, d_{i-1} + (1-a) \hat{n}_i : n_i \le d_{i-1} \end{cases}$$

where $n_i$ is the network delay of the i[th] packet, $\beta = 0.75$ and $a =$ 0.99802. The following equation is used to estimate $\hat{v}_i$ : $\hat{v}_i = a\, v_{i-1} + (1-a) \left| \hat{d}_i - n_i \right|$. At the beginning of a talkspurt, adaptive jitter buffer changes the play out delay using the equation: $D = \hat{d}_i + \mu * \hat{v}_i$ , where $D$ is the play out delay and $\mu$ is a constant that can be selected from 1 to 20. We set $\mu$ to be 4 in our simulation. It should be noted that for VoIP over wireless, the network delay $n_i$ consists of delays introduced by the wireline network and the wireless link. Jitters can be introduced by network congestions in the wireline network or by retransmissions/propagations in the wireless links. In view of the fact that most jitter buffer algorithms were proposed for compensation of network congestion jitters, it should be valuable to investigate the impact of retransmission jitters for VoIP over wireless

**III. SIMULATION SYSTEM DESCRIPTION**

Our study is based on network simulator *ns-2* [8], in which we simulated a last-hop wireless scenario. Both of the IEEE 802.11 and the Ethernet protocol stack are implemented in the simulator. A two way Bernoulli error model was inserted to simulate the wireless link transmission errors. In 802.11, if the packet size exceeds the Max. Transmission Unit (e.g. 1500 bytes for WaveLan) the packet will be fragmented. Since we set the packet size to 71 bytes, a 12.2kbit rate AMR speech frame for one RTP packet the impact of fragmentation is avoided.

The simulation system is given in Figure 1. In our simulation, the original speech file is first encoded by the AMR codec and then analyzed to extract the speech marking information (voiced/unvoiced) for each packet. The speech marking information is used with network delay and wireless link quality to control the retransmission policy. The error model determines whether a packet is corrupted or not according to

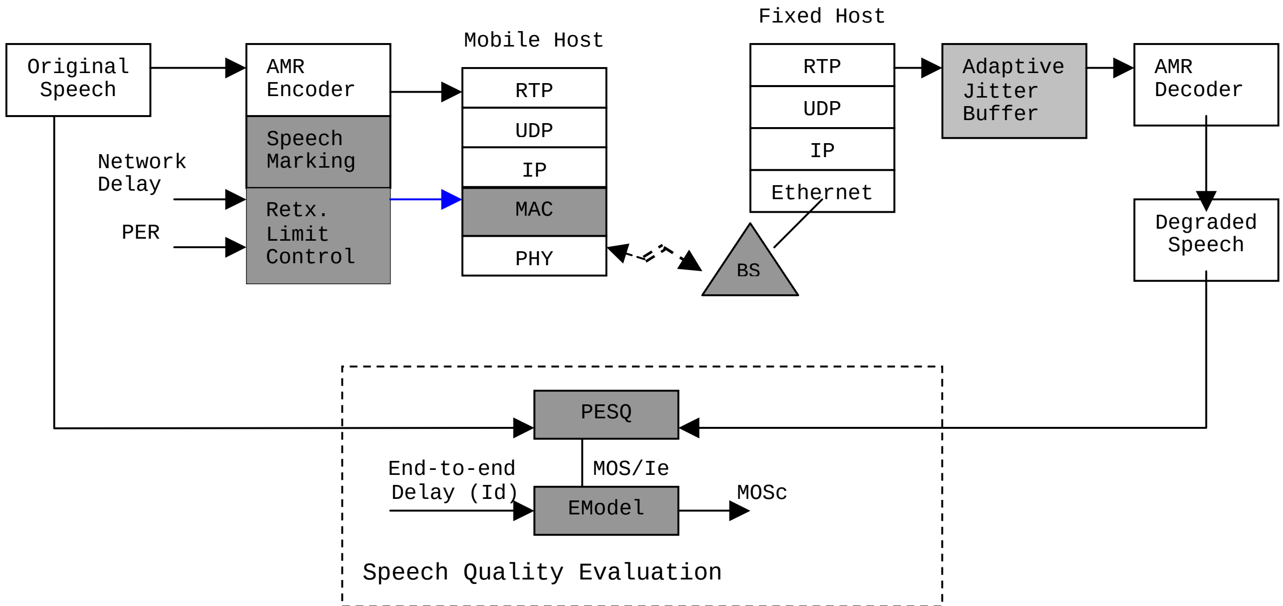


Figure 1 Simulation Environment

packet error probability ( PER). The base station (BS) will neither send an ACK to the sender for a corrupted packet nor present it to the high layer. If the MAC layer of the sender has not received an acknowledgement for a packet, it will retransmit the packet until the packet is ACKed or it reaches the limit of retransmission (we will denote Retransmission as Retx in the rest of this paper). In our simulation, we set the Retx limit to 6 for both SPB Retx and Full Retx. In the receiver, the received speech packets are fed to an adaptive jitter buffer and subsequently decoded to recover the degraded speech file that is used to obtain a measure of speech quality.

In our study, we used combined PESQ and E-Model to evaluate the conversational speech quality as described in Section II-B. Performance index was obtained averaging the computation results that were obtained from this method for each 20 seconds of the speech file.

## IV. RESULT ANALYSIS AND THE PROPOSED RETRANSMISSION SCHEME

The following simulation results were obtained by averaging results of 50 simulations with different random seeds to avoid the impact of packet loss locations. The three simulated retransmission schemes are SPB Retx, Full Retx and Null Retx.

TABLE.1 gives the average number of voiced packets losses of transmitting 73000 speech packets in our simulated wireless network with these schemes. For simplicity, we only simulated the wireless link for the purpose of this study. And only the wireless link (Retx limit exceeded) and the adaptive jitter buffer account for the packet losses. In Table.1, most of the losses of voiced packets in Full Retx or SPB Retx are caused by jitter buffer. As we deployed a Bernoulli error model in our simulation, most of the retransmitted packets can be successfully received by the receiver. If the bursty of packet errors is considered, there should be more losses of voiced packets in Full Retx or SPB Retx scheme.

TABLE.1- Average Voiced Packets Losses With *fast-exp* Jitter Buffer

| Retx Scheme / PER | No Retx | SPB Retx | Full Retx |
|---|---|---|---|
| 0.0001 | 15 | 53 | 29 |
| 0.0005 | 36 | 54 | 27 |
| 0.0008 | 61 | 51 | 26 |
| 0.001 | 69 | 47 | 22 |
| 0.003 | 144 | 28 | 17 |
| 0.005 | 241 | 22 | 13 |
| 0.01 | 474 | 13 | 9 |
| 0.05 | 2344 | 42 | 16 |
| 0.10 | 4678 | 931 | 159 |

It seems very straightforward that SPB Retx should be better than No Retx and at least the same as Full Retx with regard to the performance of protecting voiced frames. However, in TABLE.1, we can see that Full Retx always has less voiced packets losses, while No Retx has the least lost voiced packets when link quality is good (packet error probability lower than 0.0005). In fact, as in *fast-exp* algorithm, the estimated playout delay will increase with the number of retransmission jitters increases. When link quality is good, the estimated play out delay keeps at a low level, occasionally retransmitted packets and packets adjacent to them would be discarded by jitter buffer due to jitters they introduced. However, in No Retx scheme, a corrupted packet doesn’t affect its following packets. That’s why it has least packet losses when link quality is very good. On the other hand, in SPB Retx, unvoiced

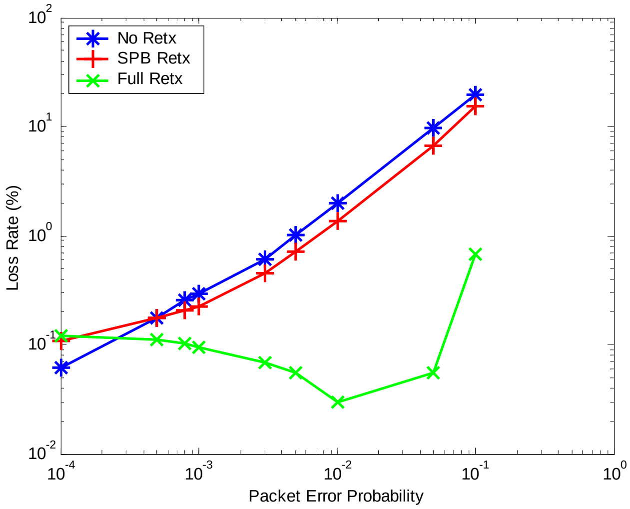


Figure 2 Overall packet loss rate comparison

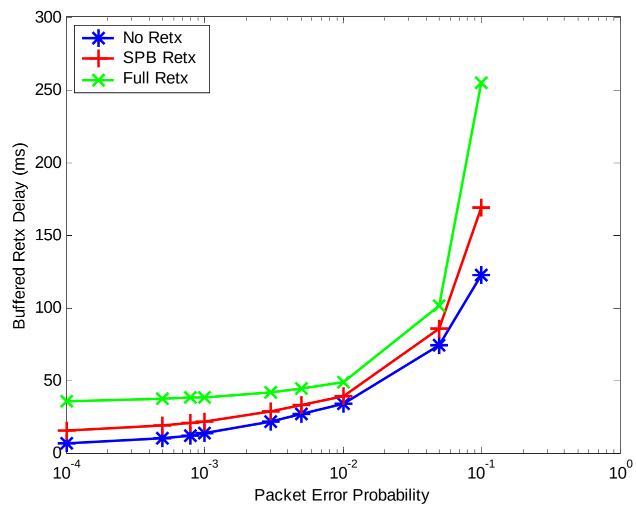


Figure 3 Buffered retx delay comparison

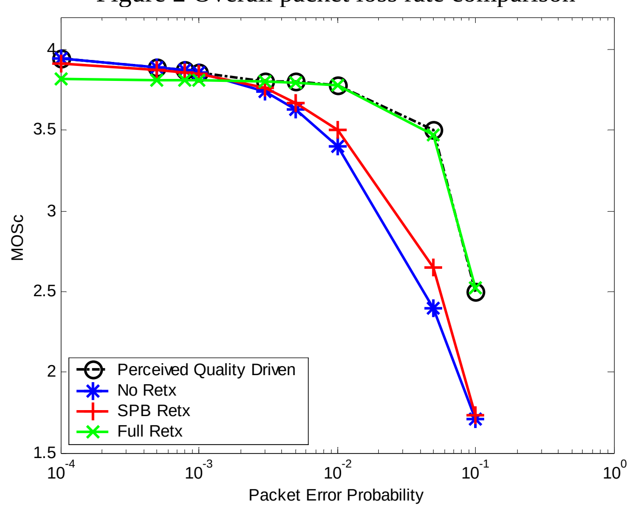


Figure 4 MOSc comparison with 175ms network

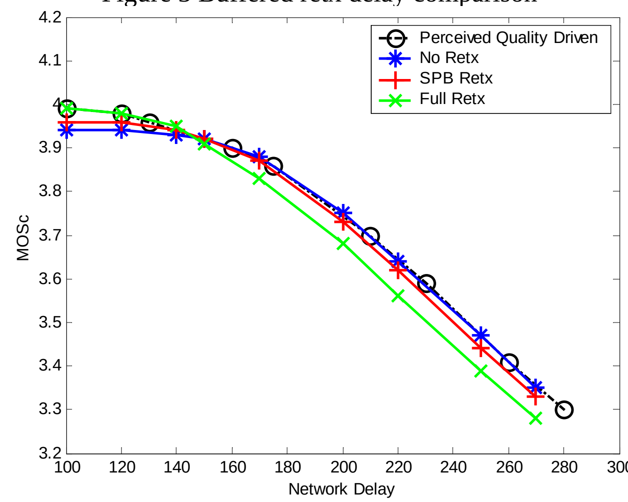


Figure 5 MOSc comparison with packet error probability 0.001

packets are not retransmitted hence the estimated playout delay can't reflect current wireless link situations when link quality becomes worse. While in Full Retx, every unACKed packets is retransmitted, this is helpful for the adaptive jitter buffer to estimate the playout delay for the next talkspurt. That's why the adaptive jitter buffer discard more packets in SPB Retx than in Full Retx.

Figure 2 and Figure 3 give the overall packet loss rates and buffered retransmission delay comparison. In Figure 2, we can see that Full Retx keeps the packet loss rate at a low level at the expense of higher delay as plotted in Figure 3 because every unACKed packet is retransmitted. It's very interesting that when link quality is not too bad (packet error probability up to 0.01), packet loss rate of Full Retx scheme is decreasing while link quality becoming worse. In fact, as we mentioned before, in worse link quality, more retransmissions helps the jitter buffer to estimate playout delay more accurately. However, when link quality is very good (packet error probability up to 0.0005), No Retx can obtain the best packet loss rate because it doesn't introduce any jitter and few packets is corrupted due to bit errors. As a compromised method, the packet loss rate and Retx delay of SPB Retx is between No Retx and Full Retx.

Using the evaluation method described in Section II-B, we give a more straightforward performance comparison in Figure 4 and Figure 5 for these schemes with MOSc as the metric. Our evaluation didn't consider the packet losses introduced in the wireline network hence to focus on the performance of Retx schemes. However, we considered network delay in the evaluation. For natural hearing, delays lower than 100ms cannot really be appreciated, but delays above 150ms can obviously affect conversation interactivity [8]. Considering Retx delays rarely exceed 100ms, to obviously reflect the impact of Retx delay, we assume 175ms delay had been introduced in the wireline network and add it to the end-to-end delay in the MOSc evaluation. In Figure 4, the MOSc of Full Retx is lower than No Retx and SPB Retx when packet error probability is lower than 0.003. That's because Full Retx scheme always introduces more Retx delay, while the perceived speech quality is sensitive to high delay when link quality is good. When packet error probability exceeds 0.003, Full Retx scheme becomes the best, as it can greatly reduce the number of corrupted packets. Fig. 5 illustrates the performance comparison with different network delays when packet error probability is 0.001. In Fig. 5, we can see that when delay lower than 150ms, Full Retx can get the best MOSc. When delay is higher than 150ms Null Retx becomes the best, it confirms that 150ms is the threshold above which delay begins to have a severe impact on speech quality. Similar to Fig 4, the performance of SPB is between No Retx and Full Retx, but it doesn't become the best in both sides of the delay threshold.

Considering both No Retx and Full Retx schemes can achieve the best MOSc under different link quality and network delay situations. We propose a new perceived speech quality driven retransmission scheme, which can switch between these two schemes when link quality and network delay changes. The pseudo code of the new scheme is shown in Figure 6. *Low_Error_Threshold* is set to be 0.0005 and *High_Error_Threshold* is 0.003. Since according the simulation results, when packet error probability is lower than 0.0005, No Retx can achieve the best MOSc even delay is not considered, whereas Full Retx becomes the best when packet error probability exceed 0.003, even network delay is very high. When packet error probability is between 0.0005 and 0.003, the decision should be made according to network delay. In the proposed scheme, *Delay_Threshold* is set to be 150ms as it's the threshold that delay begin to obviously affect speech quality. In real applications, we can convert Bit Error Rate (BER) to PER, and BER can be obtained according to bit errors in bit pattern series sent from BS. Network delay can be estimated by deducting average MH to BS handoff delay from average end-to-end delay that can be retrieved from RTP packet header.

The performance of the new perceived speech driven scheme is also given in Figure 4 and Figure 5 under different network delay and packet error probability. We can see that the curve of the perceived quality driven scheme is overlapped with parts of No Retx and Full Retx when they achieve best MOSc. As it can switch to the more suitable scheme between No Retx and Full Retx when communication conditions changes. Since this method only uses Full Retx when it's necessary, it can also achieve the similar retransmission efficiency as SPB Retx while avoid the implementation complexity to obtain speech property information that is necessary for SPB Retx.

```
if (PER < Low_Error_Threshold) .
            No_Retx();
else if (PER>High_Error_Threshold)
           Full_Retx();
else {
        if(Network_Delay<Delay_Threshold)
           Full_Retx();
        else No_Retx();
    }
```

Figure 6 Perceived speech quality driven Retx scheme pseudo code

**VII. CONCLUSION**

A suitable retransmission scheme is crucial for obtaining the best possible perceived speech quality in wireless VoIP applications. In this paper, we investigated the performance of three different retransmission schemes (No Retx, SPB Retx, Full Retx) with regard to the perceived conversational speech quality. The impact of retransmission jitters with an adaptive jitter buffer was also considered. The simulation results show that the performance of these schemes depends on the network delay and wireless link quality. Considering that the wireless environment is variable, we have proposed a perceived speech quality driven retransmission scheme that can adapt to the wireless link quality and network delay conditions. As the SPB Retx is not involved in the new method, the implementation complexity for retrieving speech property information is avoided. Our results show that the proposed method can achieve an optimum MOSc compared to No Retx, Full Retx and SPB Retx. Since the most suitable scheme is deployed by the new method when communication conditions changes. In the study, a simplified last hop wireless network is implemented to demonstrate wireless voice over IP scenario. Further improvements may be achieved by making the simulation closer to real network, e.g. by incorporating a multi-state error model in the wireless link.

**Reference:**


[1] IEEE Standards Department, 1999, IEEE 802.11 Standard for Wireless LAN, Medium Access Control (MAC) and Physical Layer (PHY) Specification.

[2] C.Hoene, I.Carreras, A.Wolisz, 2001, Voice over IP: Improving the Quality Over Wireless LAN by Adopting a Booster Mechanism – An Experiment Approach. Proc. SPIE 2001 - Voice Over IP (VoIP) Technology, pp. 157- Denver, Colorado, USA

[3] H Sanneck, N Tuong L Le et al, 2001, Selective Packet Prioritization for Wireless Voice over IP, 4th Int Sym Wireless Personal Multimedia Communication, Denmark

[4] L Sun, E.C.Ifeachor, 2003, Prediction of Perceived Conversational Speech Quality and Effects of Playout Buffer Algorithms, to appear in the Proc. of IEEE ICC 2003

[5] L.F.Sun, G.Wade, B.M.Lines and E.C.Ifeachor, 2001, Impact of Packet Loss Location on Perceived Speech Quality ,Proceedings of 2nd IP-Telephony Workshop (IPTEL '01), Columbia University, New York, pp.114-122.

[6] ITU-T G.107, The E-model, a computational model for use in transmission planning, May 2000

[7] R.Ramachandran, J.Kurose, D.Towsley and H.Schulzrinne, 1994, Adaptive playout mechanisms for packetized audio applications in wide-area networks, Proc. of IEEE Inforcom, vol.2, pp.680-688

[8] The Network Simulator - ns-2, available on line at http://www.isi.edu/nsnam/ns/

[9] ITU-T G.114, One-Way Transmission Time, Feb 1999