Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp186-191)

Optimal Adaptive Voice Smoother with Lagrangian Multiplier Method for VoIP Service

Shyh-Fang HUANG, Eric Hsiao-kuang WU and Pao-Chi CHANG
Dept of Electrical Engineering, Computer Science and Information Engineering and Communication Engineering
National Central University, Taiwan

Abstract: - VoIPs, emerging technologies, offer high-quality, real-time voice services over IP-based broadband networks. Perceived voice quality is a key metric for VoIP applications that is mainly affected by IP network impairments such as delay, jitter, and packet loss. Playout buffers at the receiving end can compensate for the effects of jitter based on a tradeoff between delay and loss. Adaptive smoothing algorithms are capable of dynamically adjusting the smoothing size by introducing a variable delay based on the network parameters to avoid the quality decay problem. This paper introduces an efficient and feasible perceived quality method for buffer optimization to achieve the best voice quality. This work formulates an online loss model that incorporates buffer sizes and applies the Lagrangian Multiplier approach to optimize the delay-loss problem. Distinct from the other optimal smoothers, the proposed optimal smoother, suitable for most codecs, carries the lowest complexity. Simulation experiments confirm that the proposed adaptive smoother achieves significant improvement in the voice quality.

Key-Words: - Adaptive voice smoother, VoIP, buffer re-synchronization, delay/loss trade off

1 Introduction

With the growing popularity of the Internet, which is traditionally used in delivering data only, novel multimedia services, such as delay bounded voice and video streaming applications are feasibly and easily delivered by broadband packet networks, such as cable modem, digital subscriber line, etc. The next generation network, like an ALL-IP network, is evolving to integrate all heterogeneous wired and wireless networks and provide seamless worldwide mobility. In an All-IP network, one revolution of the new generation Internet applications will provide VoIP services that people can talk freely around through the mobile-phones, the desktops and VoIP telephones at any time and place. Unfortunately, the IP-based network does not guarantee the available bandwidth and assure the constant delay jitters (i.e., the delay variance) for real time applications. In other words, individual transmission delays of a given flow of packets in a network may continue to change subject to varied traffic load and different routing paths caused by congestions, so that the packet network delays for a continuous series of intervals (i.e. talkspursts) at the receiver may not be the same (i.e. constant) as the sender. In addition, a packet delay may occur by the signal hand-out or the difference of bandwidth transportation in wireless/fixed networks.

The voice smooth technology usually employs jitter buffers to pre-store some voice packets for playout. A hardware device or software process that eliminates jitter caused by transmission delays in an Internet telephony (VoIP) network. For delay sensitive applications, a dominant portion of packet losses might be likely due to delay constraint. A late packet, which arrives after a delay threshold, i.e. the playback time, is treated as a lost packet. A tight delay threshold not only degrades the quality of playback but also reduces the effective bandwidth because a large fraction of delivered packets are dropped. In fact, delay and loss are normally not independent of each other. In order to reduce the loss impact, a number of applications will enlarge smoothing buffers to reduce the quality degradation caused by loss packets. However, a large buffer will induce excessive end-to-end delay and deteriorate the multimedia quality in interactive real-time applications. Therefore, a tradeoff is required between increased packet loss and buffer delay to achieve satisfactory results for playout buffer algorithms.

For perceptual-based buffer optimization schemes for VoIP, voice quality is used as the key metric because it provides a direct link to user perceived QoS. However, it requires an efficient, accurate, and objective way to optimize perceived voice quality. This paper introduces a new delay-loss smoother that employs the Lagrange multiplier method to optimize the voice quality by balancing the delay and the loss. Lagrange multiplier method is often used to optimize the trade off problems. The contributions of this paper are three-fold: (i) A new method is for optimizing voice quality for VoIP and is easily applied to new codecs. (ii) Different from the other optimal smoothers, our optimal smoother has the lowest complexity with $O(n)$. (iii) A simple scheme of the buffer resynchronization to efficiently avoid the buffer overflow. The remainder of this paper is organized as follows: In Section 2, we overview the related research works. In Section 3, we introduce the proposed novel adaptive smoother. In Section 4, the detailed description of the buffer re-synchronization solution is shown. In Section 5, the simulation results in smoothers are depicted. Finally, conclusions are drawn.

2. Related Work

The widely deployed Internet is usually lack of performance guarantee to achieve the adaptability and scalability. One of the greatest challenges to VoIP is voice quality and the keys to acceptable voice quality include bandwidth and delay. The studies of the literatures made on the degradation of the voice quality consider the effect of packet loss, but less of efforts consider that of packet delay. Recently, the research efforts about the characteristics of the end to end packet transmission delay have been initiated in some literatures [1] [2].

Within literature on predicting delays, the use of Pareto distribution in [1] is computing the distribution parameters and rebuilding the new distribution to predict the next packet delay, and the use of neural network models to learn traffic behaviors [2]. The use of Pareto distribution or a neural network model requires relatively high complexity or a long learning period. Therefore, we consider the smoothers [3]-[9] which employ statistical network parameters related with the voice characteristic, i.e. loss, delay and talkspurt that have significant influence to the voice quality. They detect delay spike in traffic and quickly calculate the required buffer size to keep the quality as good as possible.

The Spike Detection (SD) Algorithm has been studied by many researchers [3]-[9]. A delay spike is defined as a sudden and significant increase of network delay in a short period often less than one round-trip. This algorithm adjusts the smoothing size, i.e. playback delay, at the beginning of each talk-spurt. The results of this algorithm are therefore compared to the results obtained herein. The SD Algorithm in [3] used the gap-based method to detect delay spikes. For a voice session of *N* packets and *L* talkspurts, define t_i^k , a_i^k , n_k , p_i^k as the sender timestamp, receiver timestamp, number of packets, and playout time for packet *i* of talkspurt *k*. The SD Algorithm uses estimations of the mean network delay, d_i^k , and variance, v_i^k , for packet *i* of talkspurt k to adapt the playout. The mean estimation of network delay is based on the RFC 793 algorithm (see [7]), while the variance is estimated using a measure of the variation in the delays as suggested by Van Jacobson in the calculation of the round trip estimates for the TCP retransmit timer (see [3]). These estimations are recomputed each time a packet arrives, but only used when a new talkspurt is initiated. In the detection of a new talkspurt, both algorithms use the most recent values of d_i^k and v_i^k to calculate the playout time of the first packet p_i^k using Eq. (1). For all subsequent packets within the same talkspurt is used to calculate their playout time (p_i^k) .

$$p_i^k = t_i^k + d_i^k + \gamma v_i^k \tag{1}$$

where t_i^k represents the time at which packet i of talkspurt k is generated at the sending host and γ is a constant factor used to set the playout time to be “far enough” beyond the delay estimate such that only a small fraction of the arriving packets could be lost due to late arrival . The value of $\gamma = 4$ is used in simulations [3]. The estimates are recomputed each time a packet arrives, but only applied when a new talk-spurt is initiated.

The mean network delay d_i and variance v_i are calculated based on a linear recursive filter characterized by factors α and β

$$\begin{cases} If\ n_i > d_{i-1} \Rightarrow \begin{cases} d_i = \beta d_{i-1} + (1-\beta) n_i \\ v_i = \beta v_{i-1} + (1-\beta)\left|d_{i-1} - n_i\right| \end{cases} \\ If\ n_i \le d_{i-1} \Rightarrow \begin{cases} d_i = \alpha d_{i-1} + (1-\alpha) n_i \\ v_i = \alpha v_{i-1} + (1-\alpha)\left|d_{i-1} - n_i\right| \end{cases} \end{cases} \tag{2}$$

where n_i is the end-to-end delay introduced by the network and typical values of α and β are 0.998002 and 0.75 [3], respectively.

The decision to select α and β is based on the current delay condition. The condition n_i > d_{i-1} represents network congestion (*SPIKE_MODE*) and the weight β is used to emphasize the current network delay. On the other hand, n_i ≤ d_{i-1} represents network traffic is stable, and α is used to emphasize the long-term average.

In estimating the delay and variance, the SD Algorithm uses only two values α and β that are simple but may not be adequate, particularly when the traffic is unstable. For example, an under-estimated problem is when a network becomes spiked, but the delay n_i is just below the d_{i-1}, the SD Algorithm will judge the network to be stable and will not enter the *SIPKE_MODE*.

3 Optimal Smoother with Delay-Loss Trade off

The proposed optimal smoother is derived using the Lagrangian method to trade off the delay and loss. This method involves, first, building the traffic delay model and the loss model. Second, a Lagrangian cost function Q is defined using this delay and the loss models. Third, the Lagrangian cost function Q is minimized and thus the delay and loss optimized solution is obtained.

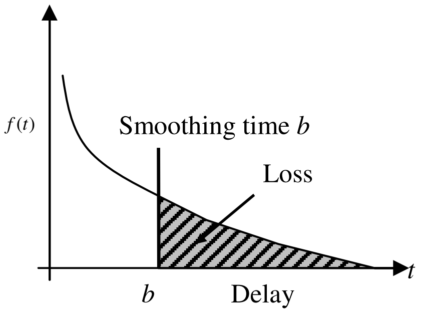


Fig. 1 The relation of smoothing delay and loss

3.1 Traffic Delay and Loss Models

For perceived buffer design, it is critical to understand the delay distribution modeling as it is directly related to buffer loss. The characteristics of packet transmission delay over Internet can be represented by statistical models which follow Exponential distribution for Internet packets (for a UDP traffic) has been shown to consistent with an Exponential distribution [10]. In order to derive an online loss model, the packet end-to-end delay is assumed as an exponential distribution with parameter $1/\mu$ at the receiving end for low complexity and easy implementation. The probability distribution function (PDF) of the delay distribution $F(t)$ can also be represented by [11][12]

$$F(t)=1-e^{tu^{-1}} \tag{3}$$

and the probability density function (pdf) of the delay distribution $f(t)$ is

$$f(t)=\frac{dF(t)}{dt}=\mu^{-1}e^{-t\mu^{-1}}\,. \tag{4}$$

In a real-time application, a packet loss that is solely caused by extra delay can be derived from the delay model $f(t)$. Figure 1 plots the delay function $f(t)$, which shows that when the packet delay exceeds the smoothing time; the delayed packet is regarded as a lost packet. The loss function $l(b)$ can be derived from Fig. 1 as

$$l(b)=\int_b^\infty f(t)dt=\left(-e^{-\mu^{-1}t}\right)\Big|_b^\infty=-e^{-\infty}+e^{-\mu^{-1}b}=e^{-\mu^{-1}b} \tag{5}$$

From Eqs. (4) and (5), we obtain the delay and loss functions that will be used in Lagrangian cost function.

3.2 Optimal Delay-Loss Adaptive Smoother

To express the corresponding quality for a given voice connection, a Lagrangian cost function Q is defined based on the delay b and the loss model $l(b)$

$$Q(b)=b+K\cdot l(b) \tag{6}$$

where $Q(b)$ represents the negative effect on voice quality, i.e., minimizing Q yields the best voice quality. K is a Lagrange multiplier where the loss becomes more significant as K increases. The K value has significant influence on the optimization process. We will discuss the valid range of the value in this section and the suggested value in the next section.

Here, once a smoothing time b is specified, the loss $l(b)=e^{-\mu^{-1}b}$ can be calculated from Eq. (5). The Lagrangian cost function in Eq. (6) yields

$$Q(b)=b+K\cdot e^{-\mu^{-1}b} \tag{7}$$

The differential equation dQ/db is assigned to zero that minimizes Q to yield the smoothing time b,

$$b=\mu\, ln\left(K\mu^{-1}\right) \tag{8}$$

where b is the best smoothing time for balancing the delay and the loss. Afterwards, the smoother can provide best quality, considering both the delay and the loss effects, based on the calculated smoothing time b.

The calculated smoothing time b is a function of K and μ. μ denotes a IP-base network delay parameter (end-to-end delay) and can be measured at the receiver, but K is given by users or applications. The calculated smoothing time b must be within an allowable range to ensure that the end-to-end delay is acceptable. Here, D_{max} is defined as the maximum acceptable end-to-end delay and the calculated smoothing time b must be between 0 and D_{max}

$$0\le ln\left(K\mu^{-1}\right)*\mu\le D_{max}\,. \tag{9}$$

Accordingly, the permissible range of valid K in the Lagrange multiplier Q function in Eq. (8) is

$$\mu\le K\le e^{D_{max}*\mu^{-1}}*\mu\,. \tag{10}$$

3.3 Suggestion of K Parameter

In this section the relationship between the voice quality and loss is further studied. Based on the previous section discussions, we know K parameter is tightly related with voice quality. In other words, for a given MOS (Mean Opinion Score) of speech quality, the allowable range of K can further be restricted. Many studies revealed the difficulty of determining the mathematical formula that relates the voice quality, delay, and loss. According to [13], the loss degrades the voice quality more remarkably than does the delay, so the quality-loss relationship is first emphasized [14][15]. In these studies, an empirical Eq. (11) was obtained by experiments with many traffic patterns for predicting the voice MOS quality MOS_{pred} that might be degraded by the traffic loss ($loss$)

$$MOS_{pred} = MOS_{opt} - c * ln(loss + 1) \qquad (11)$$

where MOS_{opt} is voice codec related, representing the optimum voice quality that the codec can achieve, c is a constant that is codec dependent, and $loss$ is a percentage ratio times 100. Following this approach, anyone can estimate a specific empirical rule with specified voice codecs and network environments. Equation (11) also implies that the network loss rate must be kept lower than or equal to the defined $loss$ to ensure the predicted MOS MOS_{pred}.

Equation (11) is rewritten to yield Eq. (12),

$$loss = 2^{\frac{MOS_{opt}-MOS_{pred}}{c}} - 1 \qquad (12)$$

Notably, the $l(t)$ function is a percentage but $loss$ is not. Therefore, $l(t)$ is multiplied by 100 to yield

$$loss = 2^{\frac{MOS_{opt}-MOS_{pred}}{c}} - 1 \ge l(t) = e^{-\mu^{-1}b} * 100 \ge 0 \quad (13)$$

From Eq. (13), the smoothing time b is

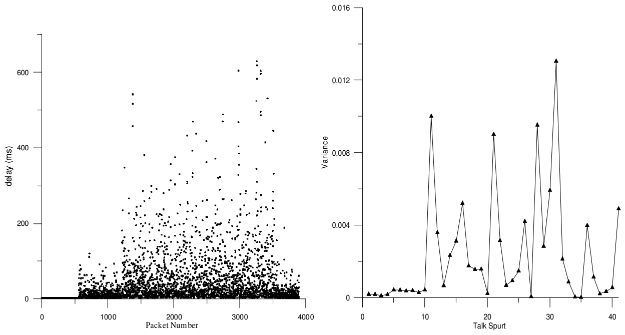
$$b \ge -ln\left(\frac{2^{\frac{MOS_{opt}-MOS_{pred}}{c}} - 1}{100}\right) * \mu . \qquad (14)$$

From Eqs. (8) and (14), the suggested range for K is

$$K \ge \frac{100 * \mu}{(2^{\frac{MOS_{opt}-MOS_{pred}}{c}} - 1)} . \qquad (15)$$

When K is assigned a value that is more than the threshold in Eq. (15), the design of the smoother is mainly dominated by the loss effect. For a given MOS, a suitable can be suggested and an optimal buffer size can be determined.

.



(a) The delay of traffic (b) The variance of traffic

Fig. 4 VoIP traffic pattern

4. Simulation

4.1 Simulation Configuration

A set of simulation experiments are performed to evaluate the effectiveness of the proposed adaptive smoothing scheme. The OPNET simulation tools are adopted to trace the voice traffic transported between two different LANs for a VoIP environment. Ninety personal computers with G.729 traffics are deployed in each LAN. The duration and frequency of the connection time of the personal computers follow Exponential distributions. Ten five-minute simulations were run to probe the backbone network delay patterns, which were used to trace the adaptive smoothers and compare the effects of the original with the adapted voice quality latter.

Fig. 3 shows the typical network topology in which a T1 (1.544 Mbps) backbone connects two LANs, and 100 Mbps lines are connected within each LAN. The propagation delay of all links is assumed to be a constant value and will be ignored (the derivative value will be zero) in the optimization process. The buffer size of the bottlenecked router is assumed to be infinite since the performance comparison of adaptive smoothers will be affected by overdue packet loss (over the deadline) and not affected by the packet loss in router buffer. The network end-to-end delay of a G.729 packet with data frame size (10 bytes) and RTP/UDP/IP headers (40 bytes) is measured for ten five-minute simulations by employing the OPNET simulation network. Table 1 summarizes the simulation parameters. Figure 4(a) and 4(b) list one of the end-to-end traffic delay patterns and the corresponding delay variances for VoIP traffic observed at a given receiver.

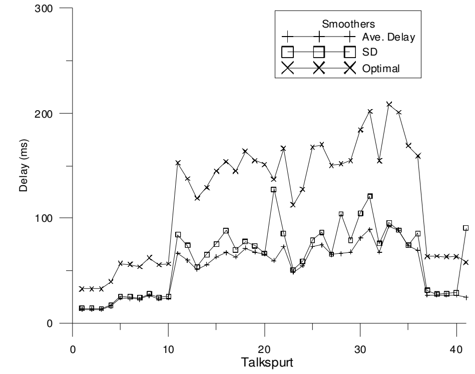


Fig. 5 The predicting time of smoothers

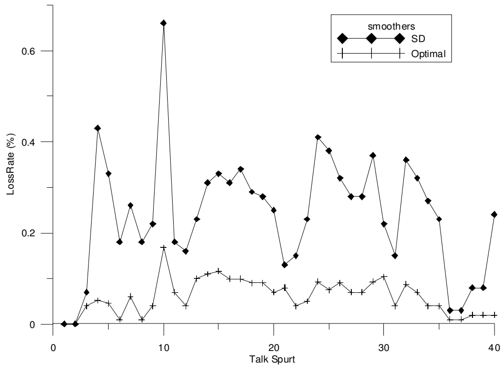


Fig. 6 The packet loss rate of smoothers

4.2 Predicted Smoothing Time and Loss Rate in Smoothers

In this section the accuracy of the predicted end-to-end delay time and loss rate among these smoothers are compared. The mean delay is used to observe the traffic pattern in particular. In Fig. 5 and Fig. 6, we can observe that the predicted time of the SD smoother is very close to the mean delay and the loss rate is higher than optimal smoother. The SD smoother uses a large value of fixed β to deal with various traffic conditions and emphasize a long-term mean delay d_{i-1}, so the predicted delay will be close to the mean delay. A better choice for n_i is probably the maximum delay in the last talkspurt that can sufficiently represent the worst case of current network congestion and avoid an under-estimated delay.

4.3 Quality Degradation with the Lagrangian Cost Function

The test sequence is sampled at 8 kHz, 23.44 seconds long, and includes English and Mandarin sentences spoken by male and female. Table 2 lists the mean delay, mean loss rate, and SSNR measured in a voice quality test with various smoothers. SSNR [16][17] is used as an evaluation tool because it correlates better with MOS and it is relatively simple to compute. Table 2 shows that the Optimal smoother performance achieves a high average SSNR and has the significant improvement in the voice quality over SD smoother, since the proposed optimal smoother truly optimizes with the delay and loss impairments. The SSNR can only represent the loss impact, but hardly represent the delay impact. Therefore, a Lagrangian cost function is utilized to consider the delay and loss impacts to the quality degradation for various smoothers. In order to maintain the normal voice quality over the network, the predicted MOS, $MOS_{pred} = 3$ is required. According to [14] and G.729, c is set as 0.25 in formula (15) and the μ is set as the frame rate 10 ms for G.729 at the sender. The Lagrange multiplier value $K = 430$ is calculated from the formula (15). Figure 7 shows the quality degradation of smoothers. From the Table 3, we can observe that the optimal smoother has the lower Lagrangian cost value than SD smoother. Specifically, we can observe the optimal smoother has 23% improvement of the quality degradation on SD smoother.

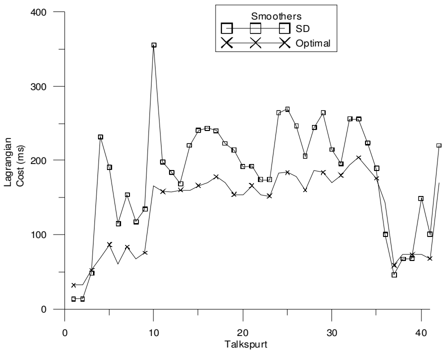


Fig. 7 The quality degradation of smoothers

4.4 Quality Score with the E-model

The E-model is a computational model, standardized by ITU-T in G.107, G.109 and G.113 [18] which uses the various transmission parameters to predict the subjective quality of packetized voice. Therefore, it is essential for the passive monitoring agent to track the performance of this channel.

In the E-model, a rating factor R represents voice quality and considers relevant transmission parameters for the considered connection. It is defined in [18] as:

$$R = Ro - Is - Id - Ie\_eff + A \qquad (17)$$

where Ro denotes the basic signal-to-noise ratio; Is denotes the sum of all impairments associated with the voice signal; Id represents the impairments due to delay of voice signals; Ie_eff denotes the equipment impairments, depending on the low bit rate codecs (Ie, Bpl) and packet loss (Ppl) levels;

Advantage factor A is no relation to all other transmission parameters.

The test sequence follows the configuration of section 5.3 and the parameters of Ro, Is and A use the default setting that were suggested by [18]. Fig. 8 shows the E-model score R of the voice quality. It shows that the optimal method has the significant improvement in the voice quality over SD smoother.

6. Conclusion

For new-generation VoIP services, a dynamic smoothing algorithm is required to address IP-based network delay and loss. This work proposes an optimal smoothing method to obtain the best voice quality by Lagrangian lost function which is a trade off between the negative effects of the delay and the loss. The buffer re-synchronization algorithm is also proposed to prevent buffer overflow by skipping some silent packets of the tail of talk-spurts. It can efficiently solve the mismatch between the capture and the playback clocks. Numerical examples have shown that our proposed method can control the playout time to balance the target delay and loss.

References:

[1] V. Brazauskas and R. Serfling, “Robust and efficient estimation of the tail index of a one-parameter pareto distribution,” North American Actuarial Journal available at http://www.utdallas.edu/~serfling, Apr. 2000.

[2] P. L. Tien and M. C. Yuang, “Intelligent voice smoother for silence-suppressed voice over internet,” IEEE JSAC, vol.17, no.1, pp.29-41, Jan. 1999.

[3] R. Ramjee, J. Kurise, D. Towsley, and H. Schulzrinne, “Adaptive playout mechanisms for packetized audio applications in wide-area networks,” Proc. IEEE INFOCOM, pp.680-686, June 1994.

[4] D. R. Jeske, W. Matragi, and B. Samadi, “Adaptive play-out algorithms for voice packets”, Proc. IEEE Conf. on Commun., vol.3, pp.775-779, 2001.

[5] J. Pinto and K. J. Christensen, “An algorithm for playout of packet voice based on adaptive adjustment of talkspurt silence periods,” Proc. IEEE Conf. on Local Computer Networks, pp.224-231, Oct. 1999.

[6] Y. J. Liang, N. Farber, and B. Girod, “Adaptive playout scheduling using time-scale modification in packet voice communications,” Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing, vol.3, pp.1445-1448, 2001.

[7] A. Kansal and A. Karandikar, “Adaptive delay estimation for low jitter audio over Internet,” IEEE GLOBECOM, vol.4, pp.2591-2595, 2001.

[8] A. K. Anandakumar, A. McCree, and E. Paksoy, “An adaptive voice playout method for VOP applications,” IEEE GLOBECOM, vol.3, pp.1637-1640, 2001.

[9] P. DeLeon and C. J. Sreenan, “An Adaptive predictor for media playout buffering,” Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing, vol.6, pp.3097-3100, 1999.

[10] J. C. Bolot, “Characterizing end-to-end packet delay and loss in the internet,” Journal of High-Speed Networks, vol. 2, pp. 305-323, Dec. 1993.

[11] F. Huebner, D. Liu and J. Fernandez, “Queueing performance comparison of traffic models for internet traffic,” IEEE GlobeCom, Sydney, vol.1, pp. 471-476 Nov. 1998.

[12] K. Fujimoto, S. Ata and M. Murata, “Statistical Analysis of Packet Delays in the Internet and its Application to Playout Control for Streaming Applications,” IEICE Trans. Commun., vol.E84-B, no.6, pp.1504-1512, June 2001

[13] K. Nobuhiko and I. Kenzo, “Pure delay effects on speech quality in telecommunications”, IEEE JSAC, vol.9, no.4, May 1991.

[14] B. Duysburgh, S. Vanhastel, B. De Vreese, C. Petrisor, and P. Demeester, “On the influence of best-effort network conditions on the perceived speech quality of VoIP connections”, Proc. Computer Communications and Networks, pp.334-339 2001.

[15] L.Yamamoto, J. Beerends, KPN Research, “Impact of network performance parameters on the end-to-end perceived quality”, EXPERT ATM Traffic Symposium available at http://www.run.montefiore.ulg.ac.be/~yamamoto/ publications.html, Sep. 1997.

[16] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, “Joint impulse response shortening for discrete multitone transceivers,” IEEE Trans..Communications, vol. 44, no. 12, pp. 1662-1672, Dec. 1996

[17] N. M. Hosny, S. H. El-Ramly, M. H. El-Said, ”Novel techniques for speech compression using wavelet transform “, The International Conference on Microelectronics, pp. 225- 229, Nov. 1999.

[18] ITU-T Recommendation G.107, “The E-model, a Computational Model for use in Transmission Planning”, Mar., 2003.