New Models for Perceived Voice Quality Prediction and their Applications in Playout Buffer Optimization for VoIP Networks

Lingfen Sun and Emmanuel Ifeachor
Centre for Signal Processing & Multimedia Communication
School of Computing, Communications and Electronics
University of Plymouth
Plymouth PL4 8AA, U.K.
Email: L.Sun@plymouth.ac.uk; E.Ifeachor@plymouth.ac.uk

***Abstract*— Perceived voice quality is an important metric in VoIP applications. The quality is mainly affected by network impairments such as delay, jitter and packet loss. Playout buffer at the receiving side can be used to compensate for the effects of jitter based on a tradeoff between delay and loss. The main aim in this paper is to find an efficient perceived quality prediction method for perceptual optimization of playout buffer. The contributions of the paper are three-fold. First, we propose an efficient new method for predicting voice quality for buffer design/optimization. The method can also be used for voice quality monitoring and for QoS control. In the method, non-linear regression models are derived for a variety of codecs (e.g. G.723.1/G.729/AMR/iLBC) with the aid of ITU PESQ and the E-model. Second, we propose the use of minimum overall impairment as a criterion for buffer optimization. This criterion is more efficient than using traditional maximum Mean Opinion Score (MOS). Third, we show that the delay characteristics of Voice over IP traffic is better characterized by a Weibull distribution than a Pareto or an Exponential distribution. Based on the new voice quality prediction model, the Weibull delay distribution model and the minimum impairment criterion, we propose a perceptual optimization buffer algorithm. Preliminary results show that the proposed algorithm can achieve the optimum perceived voice quality compared with other algorithms under all network conditions considered.**

I. Introduction

In Voice over IP (VoIP) applications, delay, jitter and packet loss are the main network impairments that affect perceived voice quality. Jitter can be partially compensated for by using a playout buffer at the receiving end, but this introduces further delay and additional packet loss. A tradeoff is necessary between increased packet loss and buffer delay to achieve satisfactory results for any playout buffer algorithm.

In the past, the choice/design of buffer algorithms was largely based on buffer delay and loss performance (e.g. a design objective could be to achieve a minimum average delay for a specified packet loss rate [1]–[3] or minimum late arrival loss [1]. This approach is inappropriate as it does not provide a direct link to perceived speech quality. From QoS perspective, the choice of the best buffer algorithm for a given situation should be determined by the likely perceived speech quality. The importance of this is now starting to be recognised [4]–[6]. For example, in [5], perceived voice quality is used to control the playout buffer in order to maximise the MOS values in terms of delay and loss. The concept of perceptual optimization has also been extended to other QoS control problems, such as joint playout buffer/FEC control [7] to maximise MOS values in terms of delay, loss and rate.

However, current methods of perceptual optimization are based on assumptions about perceived voice quality which are inappropriate. In [5], the method is based on the assumption that the effects of packet loss and delay on voice quality are linearly additive on the MOS scale which is doubtful. A further assumption is that the relationship between MOS and packet loss for codecs is linear which is not correct for most codecs. It has also been suggested in [7] that one equation may be used to represent the impairments due to packet loss for all codecs. This may not be appropriate, especially for newer codecs.

In all perceptual-based buffer design/optimisation and QoS control for VoIP, voice quality is used as the key metric because it provides a direct link to user perceived QoS. However, this requires an efficient and accurate objective way to measure perceived voice quality. Most current methods [7] [8] use the E-model [9] to predict voice quality, but the E-model requires subjective tests to derive model parameters which is time-consuming and often impractical. As a result, the E-model is only applicable to a limited number of codecs and network conditions. It is also inevitable that discontinuities exist in subjective results [10] because only a limited range of scenarios can be tested for. PESQ [11] gives a good measure of voice quality, but it is not appropriate for optimisation because of the overhead involved in its use in real-time.

In this paper, we have extended the method and developed new models which can be used for voice quality monitoring, buffer design/optimisation and for QoS control applications. As the method is based on end-to-end objective measurement instead of subjective tests, it can be easily applied to new codecs and network conditions.

For perceived buffer design, it is important 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 Normal, Exponential, Pareto and Weibull distributions depending on applications. For example, the delay distribution for Internet packets (for a UDP traffic) has been shown to be consistent with an Exponential distribution [12], whereas, Pareto distribution may be the most appropriate one to represent the tail delay characteristics for streaming media [13]. As delay characteristics may change with networks and applications, it is unclear what the appropriate delay distribution modelling is the best fit for current VoIP traffic. This motivated us to investigate the delay distribution modelling for VoIP trace data collected internationlly.

The contributions of the paper are three-fold:

(1) A new method for predicting voice quality for VoIP. In the method, a non-linear regression model is derived for each codec with the aid of the PESQ and the E-model. We illustrate the method for four modern codecs - G.729, G.723.1, AMR and iLBC. (2) Second, we propose the use of minimum impairment as a criterion for buffer optimization. This criterion is more efficient than using traditional maximum MOS score. (3) Third, we show that the delay characteristics of VoIP traffic is better characterized by a Weibull distribution than a Pareto or an Exponential distribution. Based on the new voice quality prediction model, the Weibull distribution model and the minimum impairment criterion, we propose a perceptual optimization buffer algorithm. Preliminary results show that the proposed algorithm can obtain the best voice quality when compared with other algorithms under the network conditions considered.

The remainder of the paper is structured as follows. In Section II, a new method for predicting voice quality is presented. In Section III, the perceptual optimization and minimum impairment criterion, and the delay distribution are discussed. In Section IV, a perceptual optimization buffer algorithm is proposed and the performance is compared with other algorithms. Section V concludes the paper.

II. New Models for Predicting Voice Quality

Fig 1a illustrates how the E-model may be used to predict voice quality in VoIP applications. Information about the codec, packet loss rate and delay is suitably transformed by the I_e and I_d models and then processed by the E-model to produce a MOS value. The MOS value is a prediction of what the perceived voice quality would be under these conditions. However, the I_e model is codec dependent and as indicated above, the derivation of the model parameters for each codec requires subjective tests which is impractical.

An important aim of our work is to develop an objective method which can be used to derive the I_e model for any codec without the need for subjective tests. The proposed method is depicted in Fig 1b and is based on the PESQ [11] (and the new PESQ-LQ [14]). The reference speech files are first encoded and then processed in accordance with the network impairments parameter values and then decoded to generate the degraded speech. The degraded speech and the reference speech are then processed by PESQ (or PESQ-LQ) to provide a MOS value. The MOS values can then be suitably transformed to give measured I_e values. As shown later, given a set of measured I_e values for a codec we can then derive an I_e model for the codec using regression techniques without the need for subjective tests.

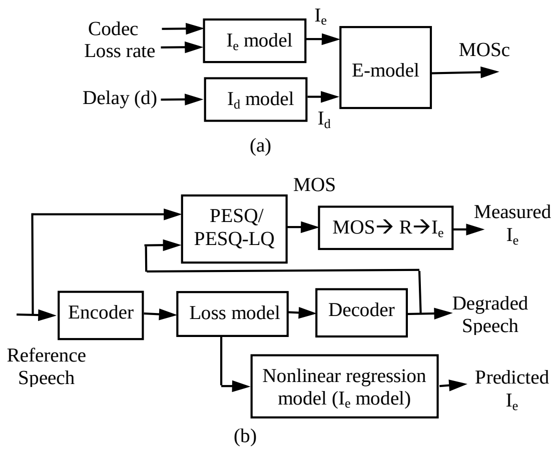

Fig. 1. (a) An illustration of how to predict voice quality using the E-model, (b) Prediction of I_e model using the PESQ

We will illustrate this for four modern codecs which are relevant for VoIP - G.729 (8 Kb/s), G.723.1 (6.3 Kb/s), AMR (the highest mode, 12.2 Kb/s and the lowest, 4.75 Kb/s) and iLBC (15.2 Kb/s). In the study, the reference speech database was taken from the ITU-T data set [15]. Packet loss was generated from 0% to 30%, in an incremental step of 3% and Bernoulli loss model was used for simplicity. PESQ-LQ (Listening Quality), the latest improvements on PESQ algorithm, is also included for comparison.

For each speech sample in the ITU-T data set, a MOS (PESQ or PESQ-LQ) score is obtained by averaging over 30 different packet loss locations (via different random seed setting) in order to remove the influence of loss location. Further, the MOS score for one loss rate is obtained by averaging over all speech samples (a total of 16 samples, consisting of 8 males and 8 females), so that the influence of gender is removed. The relationships between the average MOS and packet loss rate (expressed as ρ) for each of the four codecs are shown in Fig 2.

From Fig 2, it can be seen that PESQ-LQ has a much lower MOS score when the loss rate is high. iLBC shows the best voice quality when ρ is high (over 4%). AMR (H, 12.2 Kb/s) has the highest MOS score when ρ is zero. AMR (L, 4.75 Kb/s) has the lowest quality no matter with or without loss.

For the same data, the relationships between packet loss rate, ρ and the equipment impairment factor, I_e, for the four codecs are shown in Fig 3. The relationship between the MOS vs. ρ in Fig 2 can be converted to the Equipment impairment I_e, (measured I_e in Fig 1b) vs. ρ via Equations 1 and 2 [6].

$$R = 3.026MOS^3 - 25.314MOS^2 + 87.060MOS - 57.336 \quad (1)$$

$$I_e = R_0 - R \quad (2)$$

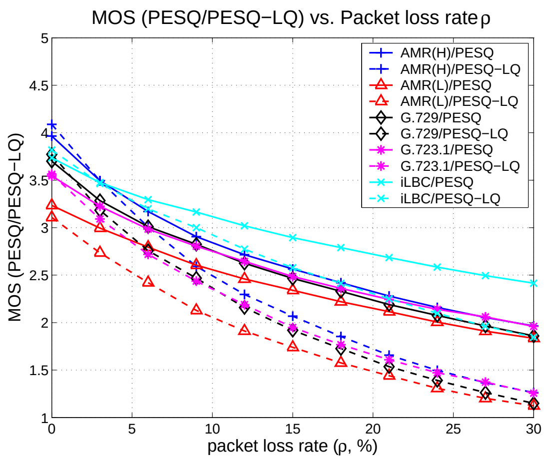

Fig. 2. MOS vs. Packet loss rate ρ

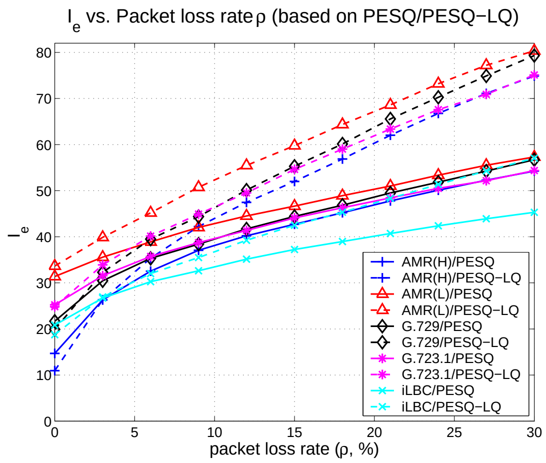

Fig. 3. I_e vs. Packet loss rate ρ

From Fig 3, a non-linear regression model (similar to the logarithm fitting function in [10]) can be derived for each codec based on the PESQ or PESQ-LQ by the least squares method and curve fitting. The derived I_e model has the following form:

$$I_e = a \ln(1 + b\rho) + c \tag{3}$$

where ρ is the packet loss rate in percentage. The parameters (a, b and c) for different codecs under PESQ and PESQ-LQ are shown in Table I and Table II, respectively.

TABLE I
PARAMETERS FOR DIFFERENT CODECS (PESQ)

Parameters	AMR (H)	AMR (L)	G.729	G.723.1	iLBC
a	16.68	30.86	21.14	20.06	12.59
b*100	30.11	4.26	12.73	10.24	9.45
c	14.96	31.66	22.45	25.63	20.42

In Fig 3, the I_e value for zero packet loss represent the codec impairment itself. The AMR (L, 4.75 Kb/s) has the largest codec impairment (the largest I_e), whereas, the AMR (H, 12.2 Kb/s) has the lowest I_e value. G.729 and iLBC codecs have similar I_e values at zero packet loss, but iLBC has the lowest I_e of all four codecs when loss rate is over 3%.

TABLE II
PARAMETERS FOR DIFFERENT CODECS (PESQ-LQ)

Parameters	AMR (H)	AMR (L)	G.729	G.723.1	iLBC
a	40.0	93.66	63.20	60.09	31.72
b*100	12.11	2.16	4.84	4.17	7.22
c	12.2	33.82	21.71	25.79	19.65

Considering that the effect of codec impairment (without loss) is fixed for any codec, I_e can be viewed as consisting of two main components: $I_e = I_{ec} + I_{e\rho}$, where I_{ec} is the impairment without loss and $I_{e\rho}$ the impairment with loss. The $I_{e\rho}$ vs. ρ for PESQ-LQ is shown in Fig 4.

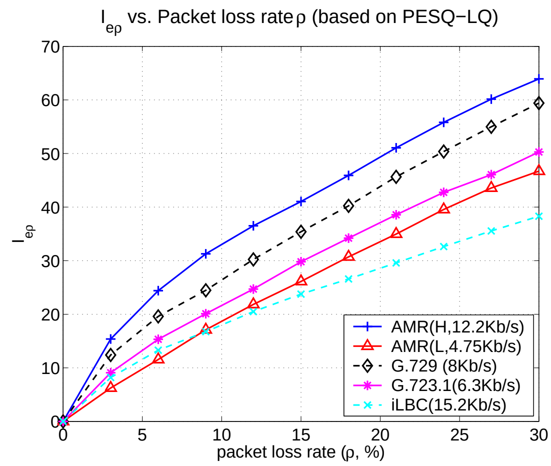

Fig. 4. $I_{e\rho}$ vs. packet loss rate ρ

Fig 4 illustrates the ability of a codec to cope with network packet loss. From the curves, the iLBC has the lowest slope, whereas, the AMR (H) has the highest. This further shows that the iLBC has an obvious high robustness to packet loss. AMR (H) has the highest MOS score under zero packet loss condition (as shown in Fig 2), but it has the least ability to cope with packet loss (quality decreases sharply as packet loss increases). From Fig 4, it is clear that to use only one curve (or model) as suggested in [7] to represent all codecs is inappropriate. Obviously with emerging new network codecs (with even higher robustness to loss), the diversity in the ability of codecs to cope with packet loss will be even larger. Thus, we recommend to use different models for each codec for accurate parameter optimization or quality control.

Unlike I_e which is codec dependent, the delay impairment factor, I_d, is common to all codecs. I_d can be derived by a simplified fitting process in [10] with Eq 4 as below.

$$I_d = 0.024d + 0.11(d - 177.3)H(d - 177.3)$$
$$where \begin{cases} H(x) = 0 \text{ if } x < 0 \\ H(x) = 1 \text{ if } x \geq 0 \end{cases} \tag{4}$$

By using I_d (Eq 4) and I_e (Eq 3), voice quality can be predicted using the E-model as shown in Fig 1(a). These models can be used for buffer optimization as described in the following section or for voice quality monitoring/control.

III. Perceptual Optimization of Playout Delay and Delay Distribution Modelling

A. Optimum voice quality and minimum impairment criterion

For perceptual buffer optimization, the aim is to achieve an optimum end-to-end voice quality (e.g. in the term of MOS score). Considering the relationship of voice quality and impairments (e.g. packet loss and delay), the problem of an optimum voice quality can be converted to an issue of minimum impairment.

We define an overall impairment function I_m which is a function of delay d and packet loss ρ, with $I_m = f(d,\rho) = I_d + I_{e\rho}$. If ignoring other impairments such as echo, R factor can be further simplified as Eq 5.

$$R = 93.2 - I_d - I_e = (93.2 - I_{ec}) - I_m \tag{5}$$

As MOS increases monotonously with R (see Eq 1), a maximum R value corresponds to a maximum MOS score. Further when maximum R is obtained, it corresponds to a minimum impairment function, I_m.

Using Eqs 3 and 4, I_m can be further expressed as:

$$\begin{aligned} I_m = I_{e\rho} + I_d = a\ln(1+b\rho)+ \\ 0.024d + 0.11(d-177.3)H(d-177.3) \end{aligned} \tag{6}$$

where a and b are codec related constants. d is the playout delay, including network delay (d_n) and buffer delay (d_b). ρ consists of network packet loss (ρ_n) and buffer loss (ρ_b).

It is a trade-off between delay and packet loss for any buffer algorithm. When playout delay $d \uparrow$ $(I_d \uparrow)$, then buffer loss $\rho_b \downarrow$ $(I_{e\rho} \downarrow)$. When $d \downarrow$ $(I_d \downarrow)$, then $\rho_b \uparrow$ $(I_{e\rho} \uparrow)$. An optimum playout delay d can be obtained when minimum impairment I_m is reached. A minimum impairment criterion for buffer optimization is set and defined in Table III.

TABLE III

Definition of a minimum impairment criterion

Given:	delay d_n, loss ρ_n and codec type
Required to estimate:	an optimized playout delay d_{opt}
Such that:	minimum I_m can be reached

Obviously seeking for a minimum I_m is more efficient than for traditionally seeking for a maximum *MOS*, as it is not necessary to convert I_m to R and then to *MOS* (a 3^{rd} order polynomial) for each buffer adaptation/calculation.

In order to find the best tradeoff of delay d and packet loss ρ, we now look at the relationship between d and ρ (or buffer loss ρ_b) which is described in the next section.

B. Playout delay and delay distribution funtion

The relationship between d and ρ_b can be described by delay Cumulative Distribution Function (CDF) which is defined as $F(x) = P(X \le x)$. For a playout delay d, the buffer loss ρ_b can be calculated as $\rho_b = P(X \ge d) = 1 - F(d)$.

To understand the delay distribution for current VoIP traffic, we investigated the delay distribution for the VoIP trace data which were collected from Internet connections between Uni. of Plymouth (UoP), UK to Beijing Uni. of Posts & Telecomm. (BUPT) China, UoP to Columbia Uni.(CU), USA, UoP to Darmstadt Uni. of Tech.(DUT), Germany, and UoP to Nanchang (NC) China. A detailed description of trace data collection is in our previous paper [6]. We experimented with Exponential, Pareto and Weibull distributions. The definition of CDF for three distributions are listed in Table IV. The RMSE (Root Mean Square Error) for the five selected traces for different approximation models are tabulated in Table V. The empirical and fitted CDF for trace 1 is illustrated in Fig 5.

TABLE IV

Definition of several cumulative probability distributions

Distribution	Exponential	Pareto	Weibull
CDF: F(x)	$1 - e^{-(x-\mu/\beta)}$	$1-(k/x)^{\alpha}$	$1-e^{-((x-\mu)/\alpha)^{\gamma}}$

TABLE V

RMSE of different distribution functions for different traces

Traces	Exp.	Pareto	Weibull
1 (BUPT $\rightarrow$ UoP, 7/6/02)	0.04467	0.03916	0.005607
2 (UoP $\rightarrow$ CU, 3/04/02)	0.0007858	0.0007389	0.0007233
3 (UoP $\rightarrow$ BUPT, 11/06/02)	0.05228	0.03398	0.01064
4 (UoP $\rightarrow$ DUT, 10/06/02)	0.01926	0.02029	0.004269
5 (UoP $\rightarrow$ NCT, 30/05/02)	0.01376	0.01366	0.003032

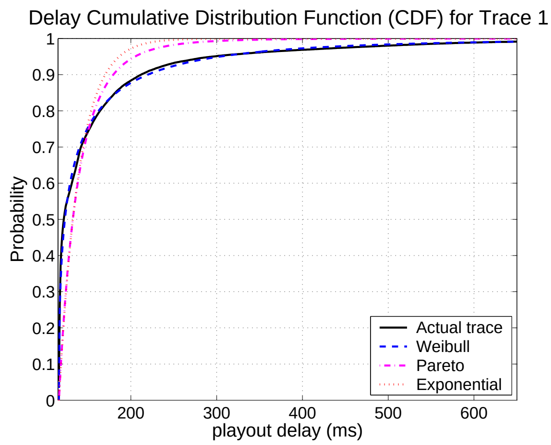

Fig. 5. Empirical and fitted CDF for trace 1 (Weibull: μ = 116, α = 15.9, γ = 0.4451; Pareto: k = 116, α = 5.277; Exp: μ = 116, β = 23.47)

From Table V, Fig 5, it can be seen that Weibull distribution achieved the best fit for all five traces (with the lowest RMSE) when compared with Pareto and Exponential distribution. As a result, we use Weibull distribution to represent delay distribution in the perceptual-based buffer design.

C. Perceptual Optimization of Playout Delay

Given network packet loss ρ_n (in percentage) and playout delay d, the buffer loss (ρ_b) for a Weibull Distribution can be calculated in the following Equation.

$$\rho_b = (1 - \rho_n/100)P(X \geq d) = (1 - \rho_n/100)e^{-((d-\mu)/\alpha)^\gamma} \tag{7}$$

Replacing ρ_b of Eq 7 into Eq 6, overall impairment factor, I_m, can be depicted as follows:

$$\begin{aligned} I_m &= 0.024d + 0.11(d - 177.3)H(d - 177.3) + \\ &a \ln\left[1 + b[\rho_n + (100 - \rho_n)e^{-((d-\mu)/\alpha)^\gamma}]\right] \end{aligned} \tag{8}$$

For a given trace segment, the Weibull Distribution location parameter μ equals to the minimum network delay d_n, the scale parameter α and shape parameter γ can be estimated using maximum-likelihood-estimator (MLE) method [16] (we use Matlab's $weibfit$ function directly in the simulation for simplicity). The optimum playout delay (d_{opt}) can be obtained by searching for a playout delay d which meets the minimum impairment criterion. Fig 6 shows an example of I_m vs. d for a trace segment (with 1000 packets) selected from trace #1. In order to see how different codecs and objective measurement methods (e.g. PESQ/PESQ-LQ) affect playout delay optimization, Fig 6 also shows I_m vs. d for AMR122 and iLBC using PESQ and PESQ-LQ. It is obviously that the optimum playout delay differs according to which codec and which objective quality method are used. The iLBC/PESQ has the smallest optimum playout delay (d_1) and AMR122/PESQ-LQ has the largest one (d_4). The minimum impairment values obtained also differ for different codecs, with iLBC (PESQ) the lowest I_m and AMR122 (PESQ-LQ) the highest I_m.

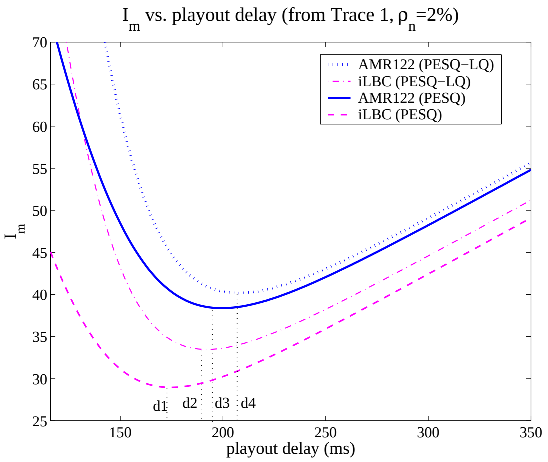

Fig. 6. Optimization of playout delay

IV. Perceptual Optimization Buffer Algorithm

A. Perceptual Optimization Buffer Algorithm (P-optimum)

In Section III, we have derived Eq 6 which relates impairment (I_m) with playout delay (d) and network packet loss (ρ_n) for a given trace. This can be used directly for perceived jitter buffer algorithm optimization. For simplicity, we only use the equation for G.723.1 codec to show the concept of perceptual optimization buffer design.

As network traces show high possibility of "spike" which is defined as a number of packets that have significantly higher delays than the rest. The "spike" state can be regarded as an exceptional state in the trace data (seen as a short-term delay characteristics) and the remaining "non-spike" state can be analysed in long-term delay distribution. Several algorithms exist for spike detection. For example, Ramachandran et al [1] proposed to use $(n_i - n_{i-1}) > threshold$ as the detection of a start of a spike (n_i is the network delay for i^{th} packet). This accounts for the spike with a sudden increase of delay. However through the analysis of our collected Internet trace data, we notice that large amounts of spike is with gradual increase which cannot be detected by the above algorithm. Considering spikes with sudden or gradual increase, we follow the spike detection based on ($n_i > threshold$) as in [2]. The proposed perceptual optimum buffer algorithm (P-optimum) is shown in Algorithm 1.

Algorithm 1 Perceptual Optimum Buffer Algorithm

For every packet i received, calculate the network delay n_i
if $mode$ == SPIKE **then**
  if $n_i \leq tail \times old\_d$ **then**
    $mode$ = NORMAL /* the end of a spike */
  end if
else if $n_i > head \times d_i$ **then**
  $mode$ = SPIKE /* the beginning of a spike */
  /* save d_i to detect the end of a spike later */
  $old\_d = d_i$
else
  /* normal model*/
  - update delay records for the past W packets
end if

At the beginning of a talkspurt
if $mode$ == SPIKE **then**
  $d_i = n_i$ /* estimated playout delay d_i */
else
  - obtain (μ, α, γ) in Weibull distribution
  - search playout delay d for $d_i = d_{opt}$ which meets $\rightarrow$ $min(I_m)$
end if

Depending on the current mode, the playout delay for the next talkspurt is estimated differently in each mode as shown in Algorithm 1. In spike-detection mode, the delay of the first packet of a talkspurt becomes the estimated playout delay for the talkspurt. Otherwise, the perceptually optimized playout delay based on the delay distribution of the last W packets (in *NORMAL* mode) is used. The large the W value, the less responsive the scheme to adapt. The *head* and *tail* parameters are used to set the threshold for spike detection.

B. Performance Analysis and Comparison

In order to compare with other buffer algorithms, we also implemented "exp-avg", "fast-exp", "min-delay", "spk-delay" and "adaptive" algorithms (detail see [6]). The results are shown in Table VI for the above five traces. The window size W is set to 1000. The *head* is 4 and the *tail* is 2, as suggested in [2]. During the experiment, we changed the window size W from 100 packets (3sec) to 10,000 packets (300 sec, as suggested by [2] and [5]), we noticed that the performance (the overall MOS score) does not show a big difference within the range. We chose W of 1000 (30 sec), as it is an appropriate duration for the I_m or MOS calculation and has higher computation efficiency than the longer window length.

From Table VI, it can be seen that "P-optimum" obtained the optimum MOS scores among all the five traces. Our previous proposed "adaptive" algorithm achieved sub-optimum results. The remaining buffer algorithms can achieve good results only in some traces, but not for all. It has to be mentioned that P-optimum has the highest complexity, whereas the others including "adaptive" have the similar low complexity.

TABLE VI
PERFORMANCE COMPARISON FOR DIFFERENT BUFFER ALGORITHMS

Trace	Buffer algorithms	Loss ρ (%)	Delay d (ms)	MOS
Trace 1	Exp-avg	4.9	298.5	2.01
	Fast-exp	1.5	750.8	1.00
	Min-delay	9.4	208.8	2.34
	Spk-delay	10.4	225.0	2.18
	Adaptive	9.0	208.1	2.37
	P-optimum	10.5	188.2	2.43
Trace 2	Exp-avg	1.8	27.3	3.28
	Fast-exp	0	35.9	3.44
	Min-delay	1.7	27.3	3.29
	Spk-delay	3.4	24.9	3.15
	Adaptive	0	35.9	3.44
	P-optimum	0.1	44.5	3.42
Trace 3	Exp-avg	18.2	432.4	1.01
	Fast-exp	14.3	1408.6	1.00
	Min-delay	22.1	312.7	1.30
	Spk-delay	23.8	325.4	1.22
	Adaptive	22.1	299.8	1.35
	P-optimum	32.0	171.1	1.80
Trace 4	Exp-avg	5.9	24.0	2.97
	Fast-exp	4.3	94.4	2.99
	Min-delay	5.3	23.0	3.01
	Spk-delay	7.6	21.9	2.86
	Adaptive	4.3	72.8	3.02
	P-optimum	5.1	34.4	3.02
Trace 5	Exp-avg	3.5	150.9	2.98
	Fast-exp	0.5	176.1	3.22
	Min-delay	4.5	148.8	2.91
	Spk-delay	6.3	144.3	2.79
	Adaptive	0.5	170.3	3.22
	P-optimum	0.5	169.8	3.22

V. CONCLUSIONS

In this paper, we have proposed a non-linear regression model to predict perceived voice quality based on PESQ/PESQ-LQ and E-model. We derived new models for variety of codecs for VoIP applications. These models can be efficiently used for voice quality monitoring, perceptual buffer design/optimization and other QoS control purposes. As the method is based on objective tests instead of subjective tests, it can be easily extended to other new codecs or network conditions. We proposed the use of minimum overall impairment as a criterion for quality control and buffer optimization. This is more efficient than traditional maximum MOS score criterion. We investigated delay distribution characteristics based on VoIP trace data collected. We find that a Weibull distribution is a better fit than a Pareto and Exponential distribution. Based on the derived voice quality prediction models, the Weibull delay distribution model and the minimum impairment criterion, we proposed a perceptual optimization playout buffer algorithm. Preliminary results show that the proposed algorithm can achieve the optimum perceived voice quality compared with other algorithms under all network conditions considered.

As the work is based on the buffer adaptation at the beginning of each talkspurt, it cannot adapt to any delay changes during a talkspurt. Future work will extend the idea to consider buffer adaptation during a talkspurt in order to achieve a best trade-off among delay, loss and end-to-end jitter.

ACKNOWLEDGEMENT

The work is supported in part by an EU grant under the Sixth Framework Programme (BIOPATTERN Project, No. 508803) and by Acterna.

REFERENCES

[1] R. Ramachandran, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide-area networks," *Proc. of IEEE Infocom*, vol. 2, pp. 680–688, 1994.
[2] S. B. Moon, J. Kurose, and D. Towsley, "Packet audio playout delay adjustment: performance bounds and algorithms," *Multimedia Systems*, vol. 6, pp. 17–28, 1998.
[3] V. Ramos, C. Barakat, and E. Altman, "A moving average predictor for playout delay control in VoIP," in *Proc. Quality of Service - IWQoS 2003, 11th International Workshop*, pp. 155–173, June 2003.
[4] A. P. Markopoulou, F. A. Tobagi, and M. Karam, "Assessment of VoIP quality over Internet backbones," *Proc. of IEEE Infocom*, 2002.
[5] K. Fujimoto, S. Ata, and M. Murata, "Adaptive playout buffer algorithm for enhancing perceived quality of streaming applications," *Proceedings of IEEE Globecom2002*, Nov 2002.
[6] L. Sun and E. Ifeachor, "Prediction of perceived conversational speech quality and effects of playout buffer algorithms," in *Proceedings of IEEE ICC'03*, pp. 1–6, 2003.
[7] C. Boutremans and J. Y. Le Boudec, "Adaptive joint playout buffer and FEC adjustment for Internet telephony," in *Proceedings of IEEE INFOCOM'2003*, pp. 652–662, April 2003.
[8] M. Gardner, V. S. Frost, and D. W. Petr, "Using optimization to achieve efficient quality of service in voice over IP networks," in *Proceedings of IPCCC 2003*, April 2003.
[9] International Telecommunicaion Union, "The E-model, a computational model for use in transmission planning," *ITU-T Recommendation G.107*, July 2000.
[10] R. G. Cole and J. Rosenbluth, "Voice over IP performance monitoring," *Journal on Computer Communications Review*, vol. 31, April 2001.
[11] International Telecommunication Union, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *ITU-T Recommendation P.862*, Feb 2001.
[12] 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.
[13] K. Fujimoto, S. Ata, and M. Murata, "Playout control for streaming applications by statistical delay analysis," *Proceedings of IEEE ICC*, vol. 8, pp. 2337–2342, June 2001.
[14] A. W. Rix, "Comparison between subjective listening quality and p.862 PESQ score," *Proceedings of Online Workshop Measurement of Speech and Audio Quality in Networks*, pp. 17–25, May 2003.
[15] International Telecommunicaion Union, "Objective measuring apparatus, Appendix 1: Test signals," *ITU-T Recommendation P.50*, Feb 1998.
[16] A. Feldmann, "Characteristics of TCP connection arrivals," *Technical report, AT&T Labs Research*, 1998. `http://citeseer.nj.nec.com/feldmann98characteristics.html`.