2005-06-01

Assessing the Quality of VoIP Transmission Affected by Playout Buffer Scheme

Miroslaw Narbutt
Technological University Dublin, miroslaw.narbutt@tudublin.ie

Mark Davis
Technological University Dublin, mark.davis@tudublin.ie

Funder: Science Foundation Ireland (03/IN3/1396)

Assessing the quality of VoIP transmission affected by playout buffer scheme

Miroslaw Narbutt, Mark Davis

Communications Network Research Institute, Dublin Institute of Technology
Phone: +353 1 402 790, Fax: +353 1 402 7901,

narbutt@cnri.dit.ie, mark.davis@cnri.dit.ie

Abstract

Delay, echo, encoding scheme, and packet loss all influence perceived quality of conversational speech transmitted over packet networks. Therefore, the choice of a buffer algorithm cannot be solely based on statistical loss/delay trade-off metrics. Also subjective "listening tests" or the newer ITU-T PESQ method, which don't consider the effect of mouth-to-ear delay are inappropriate. We proposed a method for assessing VoIP call quality by extending the ITU-T E-model concept. This method provides a direct link to perceived conversational speech quality by estimating user satisfaction from the combined effect of information loss, delay and echo.

Keywords

jitter, adaptive playout, E-model user satisfaction,

1. Introduction

Traditional subjective "listening tests" and newer objective measurement methods such as PESQ do not take into account delay impairments and therefore cannot be used to assess the end-to-end conversational call quality. Such methods are typically used by speech codec designers for assessing narrow-band speech quality and not by network planners that must deal with delay sensitive VoIP transmission.

Perceptual quality assessment has to take into account the complete end-to-end transmission that depends largely on the playout buffer scheme implemented at the receiver.

Currently the management of the playout buffer is not specified by any standard and is vendor specific. As a result there are many different adaptive and fixed playout schemes (each with a different parameter set) to chose from.

Given that information on the implementation of the playout buffer in commercial applications is practically nonexistent (the playout buffer module has a strategic value from the vendor's perspective and is usually kept confidential) there is a need for a method to evaluate buffering strategies in a VoIP system.

In this paper we present a new approach on how to assess the impact of the playout machanism implemented at the receiver on the quality of VoIP transmission. This method extends the ITU-T E-model concept and provides a direct link to the perceived conversational speech quality by estimating user satisfaction from the combined effects of information loss (due to encoding scheme and packet loss), delay and echo.

The remaining sections of the paper are structured as follows: Section 2 focuses on adaptive anti-jitter buffering and the fundamental trade-off that exists between buffering delay and packet loss due to late packet arrival. The new method for assessing user satisfaction, which extends the ITU-T E-model methodology, is described in section 3. Experimental results and their analysis are given in section 4. Section 5 concludes the paper.

2. Adaptive buffering for jitter compensation

Large delay variations in IP networks complicate the proper reconstruction of the speech signal at the receiver. To compensate for jitter a typical VoIP application buffers incoming packets in the de-jitter buffer before playing them out. This allows slower packets to arrive on time in order to be played out at the rate they were generated at the sender. Buffering delay cannot be too short or too long. If the buffering delay is too short, "slower" packets will not arrive before their designated playout time and voice quality suffers. If the buffering delay is too long, it noticeably disrupts interactive conversational speech communication.

The problem of transforming network layer delay variations to application layer loss and delay is addressed in the new ITU-T Recommendation G.1020 [1]. Packets that arrive with various impairments (delays, jitter, and errors) are processed by the application that transforms jitter into other impairments i.e. packet loss and additional delay by means of de-jitter buffering as shown in Figure 2. Packets with delay variation in the "white" range are accommodated, while packets with greater delay variation (in the "black" range) are discarded. In this way transport layer delay variation can be mapped to application layer delay and packet loss.

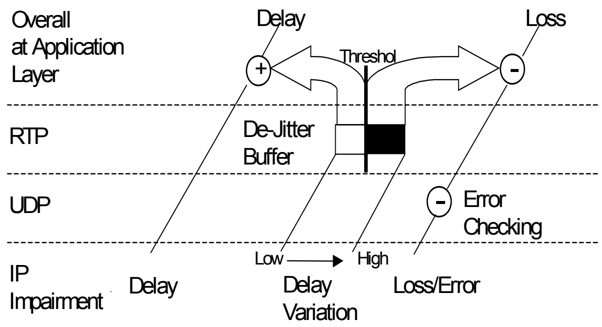


Fig. 1 Mapping IP packet performance to application layer [1]

In order to compensate for jitter the optimal delay for the de-jitter buffer should be equal to the total variable delay along the connection. Unfortunately it is not possible to find an optimal, fixed de-jitter buffer size when network conditions vary in time. Therefore, de-jitter buffers with dynamic size allocation, so called "adaptive playout buffers", are more appropriate. A good de-jitter buffer should keep the buffering time as small as possible while minimizing the number of voice packets that arrive too late to be played out. These two conflicting goals have led to various playout algorithms that calculate playout deadlines. A typical playout buffer algorithm monitors the time-stamp t_i and reception time of the *i-th* packet and adjusts the playout delay p_i accordingly as shown in Figure 2.

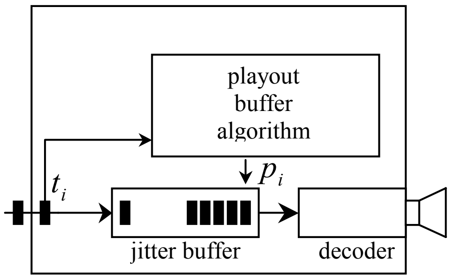


Fig. 2 Adaptive playout buffer control mechanism.

The adjustment of playout delay is achieved usually by compressing or expanding silent periods between consecutive talkspurts [2][3][4][5]. With this "per talkspurt" mechanism, the playout time is calculated only for the first packet of the incoming talkspurt. Any subsequent packets of that talkspurt are played out with the rate equal to the generation rate at the sender. This mechanism uses the same playout delay throughout a given talkspurt but permits different playout delays for different talkspurts. The variation of the playout delay introduces artificially elongated or reduced silent periods between two consecutive talkspurts.

The effectiveness of the "per talkspurt" mechanisms is limited when talkspurts are long and network delay variation is high within them. Therefore some algorithms adjust the playout time of voice packets on a "per-packet" basis. In this "per-packet" mechanism, proper reconstruction of continuous output speech is achieved by scaling individual voice packets using the "time-scale modification technique" [6]. This technique modifies the rate of playout while preserving the pitch. According to [6], voice packet can be scaled to 50% - 200% of its original size without degrading the sound quality. Authors claim that playout time adaptation within talk spurt provides the best performance in terms of loss rate and buffering delay.

A fundamental trade-off exists between buffering delay and packet loss. This trade-off is determined by the size of the de-jitter buffer. A larger de-jitter buffer can accommodate packets with greater delay variation; hence fewer packets would be lost, at the expense of larger overall delay. Similarly, a smaller de-jitter buffer will produce less overall delay, but cause a larger fraction of packets to be discarded by the terminal, thus increasing the overall loss.

Generally, a good playout algorithm should be able to minimize both: buffering time and late packet loss and thus improve the loss/delay trade-off. The loss/delay trade-off curve for a given playout algorithm can be obtained by considering average buffering delays and late loss percentages for the entire range of values of its control parameter. According to [5] the average buffering delay and late packet loss are calculated only for the accommodated packets (those that arrived before their playout deadlines). Once the loss/delay trade-off curves are obtained, it is possible to judge which algorithm performs better. If a loss/delay curve achieved by one algorithm lies below the curve achieved by a second algorithm, then the first algorithm performs better. This is illustrated in Figure 3.

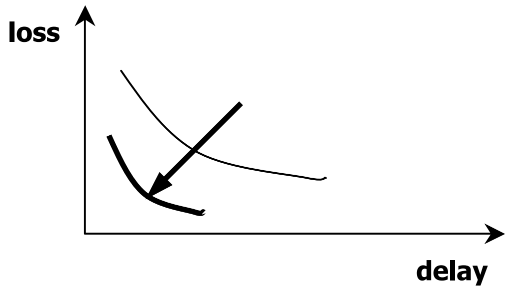


Fig. 3 Improving the trade-off between buffering delay and late packet loss.

Although this loss/delay trade-off is useful, a more informed choice of buffer algorithm can be made by considering its effect on perceived speech quality.

3. The new method for assessing user satisfaction

3.1 Related work

3.1.1 Subjective testing and objective measurements

Several voice quality assessment methods are in use and are described in different recommendations. One of the most common methods is to perform laboratory tests (e.g. "listening only tests"), where the test subjects are requested to classify the perceived quality into categories. Traditionally, perceived voice quality is defined according to the 5-grade scale known as “mean listening-quality opinion score”, or simply “Mean Opinion Score”, (MOS). An assessment of the speech transmission quality can also be obtained by calculating the percentage of all test persons rating the configuration as "Good or Better" or as "Poor or Worse". For a given connection these results are expressed as "Percentage Good or Better" (GoB) and "Percentage Poor or Worse" (PoW). A detailed description of the method, and the MOS, GoB, and PoW ratings can be found in the ITU-T rec. P.800 [7].

Subjective testing is considered as the most “authentic” method of measuring voice quality. On the other hand it is time consuming and a costly process. Moreover, subjective listening tests are very difficult to repeat and never give identical results. In contrast to subjective tests, objective testing methods are used to analyse the distortion that has occurred on test voice signals transmitted through a VoIP network. An estimate of the audible error is derived by subtracting an examined and a reference voice signal and mapping the result to the MOS scale. This testing technique called “Perceptual Speech Quality Measure” (PSQM) was recommended by the ITU-T in 1996 as P.861 [8] to assess speech codecs, used primarily for mobile transmission, such as GSM. Recognized as having certain limitations in specific areas of application PSQM was withdrawn from the ITU-T set of standards and replaced by the newer method called “Perceptual Evaluation of Speech Quality” (PESQ) in 2001. This newer method, described in ITU-T standard P.862 [9] contains an improved objective speech quality assessment algorithm. PESQ is designed for one-way listening-only perceived quality measurement and requires a reference signal. The most useful result of PESQ is the MOS that directly expresses the voice quality. The PESQ MOS as defined by the ITU recommendation P.862 ranges from 1.0 (worst) up to 4.5 (best).

3.1.2 The E-model

Subjective “listening tests” and objective measurement methods such as PESQ do not take into account delay impairments and therefore cannot be used to assess the perceived conversational speech quality. A tool that can be used to predict subjective quality of a conversational speech quality is the ITU-T E-model. The E-Model was originally developed by ETSI [10] as a transmission planning tool, and then standardized by the ITU as G.107 [11] and suggested by TIA [12] as “a tool that can estimate the end-to-end voice quality, taking the IP telephony parameters and impairments into account”. This method combines individual impairments (loss, delay, echo, codec type, noise, etc.) due to both the signal’s properties and the network characteristics into a single R-rating. The transmission rating factor R can lie in the range from 0 to 100: high values of R in a range of $90 < R < 100$ should be interpreted as excellent quality, while a lower value of R indicates a lower quality. Everything below 50 is clearly unacceptable and everything above 94.15 is unobtainable in narrowband telephony. The primary output of the E-model is the transmission rating factor R. Based on this factor, one can easily predict how an “average user” would rate a VoIP call using subjective MOS scores. The relationship between R-rating and MOS scores is depicted in Figure 4.

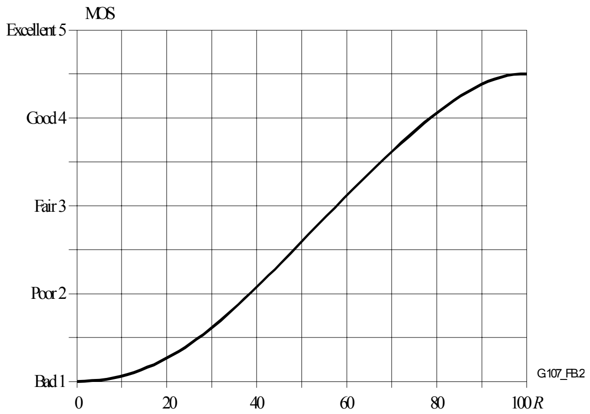


Fig. 4: MOS as function of rating factor R [10]

Based on the transmission rating factor R, ITU-T Recommendation G.109 [13] also introduces categories of user satisfaction. The definitions of those categories in terms of ranges of R are found in Table 2. Also provided is the relation between R and the MOS score.

R	MOS	User satisfaction
90 – 94.5	4.34 – 4.50	very satisfied
80 – 90	4.03 – 4.34	satisfied
70 – 80	3.60 – 4.03	some users dissatisfied
60 – 70	3.10 – 3.60	many users dissatisfied
50 – 60	2.58 – 3.10	nearly all users dissatisfied
0 – 50	1.00 – 2.58	not recommended

Table 1: Definition of categories of user satisfaction [13]

The R-rating is defined as a linear combination of the individual impairments and is given by the following formula [11]:

$$R = (R_o - I_s) - I_d - I_e + A \qquad (1)$$

where:

- R_o - Basic signal-to-noise ratio which represents subjective quality impairment due to circuit noise, room noise at sending and receiving sides, and subscriber line noise (max value R_o = 94.15 for narrowband telephony speech);
- I_s - Simultaneous impairment factor which represents subjective quality impairments due to loudness, side tone, and quantization distortion;
- I_d - Delay impairment factor which represents subjective quality impairments due to talker echo, listener echo, and absolute delay;
- I_e - Equipment impairment factor which represents subjective quality impairments due to information loss (caused by low bit rate speech coding and packet loss);
- A - Advantage factor which represents the effect of the convenience of mobile or other communication on a subjective quality.

In the context of this work, delay impairment I_d and equipment impairment I_e are the most interesting. Other impairments: loud connection and quantization impairment I_S, basic signal to noise ratio R_0, and the "advantage factor" A do not depend on the transmission parameters [14]. Therefore, we can conclude that we can write the R rating (for undistorted G.711 audio) as:

$$R = 94.15 - I_d - I_e \qquad (2)$$

Delay Impairment Factor I_d

Mouth-to-ear delay is defined in the E-model as the time between the speaker making an utterance and the moment the listener hears it. In order to preserve an acceptable level of conversation interactivity, this delay should be kept below a defined bound.

The generally-accepted limit for high-quality voice connection delay is 150 ms and 400 ms as a maximum tolerable limit. If the mouth-to-ear delay exceeds defined bounds it noticeably disrupts interactive communication. As delays rise over this figure, talkers and listeners become un-synchronized, and often they speak at the same time, or both wait for the other to speak. This condition is commonly called, talker overlap. Even if overall speech quality is acceptable, holding such a conversation can be annoying. ITU-T recommendation G.114 [15] gives the following conclusions:

- small delays (10-15 ms) are not annoying for users and no echo cancellation is required.
- delays up to 150 ms require echo control but do not compromise the effective interaction between users
- if the delays are in the range 200 ms to 400 ms, the effectiveness of the interaction is lower but can be still acceptable
- if the delay is higher than 400 ms, interactive voice communication is difficult or impossible and conversational rules are required (as “over” indicators)

Talker and listener echo both contribute significantly to perceived speech quality in VoIP telephony. As a general rule, the perceived quality decreases with increasing delay and/or increasing level of the received echo signal but listener echo can be neglected if there is sufficient control of the talker echo. The degree of annoyance of talker echo depends on the level difference between the original voice and the received echo signal. This level difference is characterized by so-called “Talker Echo Loudness Rating” (TELR). ITU-T Recommendation G.131 provides useful information regarding talker echo as a parameter by itself [16].

The relation between the mean one-way delay and the E-model rating factor R for three values of TELR is shown in Figure 5 [12].

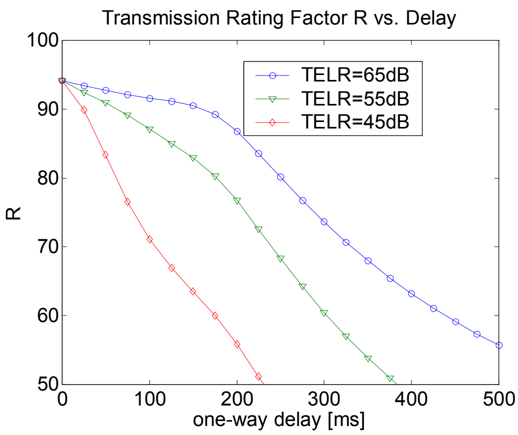


Fig. 5: Transmission rating factor R as a function of the one-way delay [12].

Equipment impairment factor I_e

Equipment impairment factor *Ie* captures effects of information loss, due to both encoding scheme and packet loss (including late packet arrival). ITU-T Recommendation G.113 [17] gives detailed values of this impairment factor for various codecs as a function of packet loss.

Figure 17 show for several codecs (and PLC techniques) how the equipment impairment increases as packet loss increases.

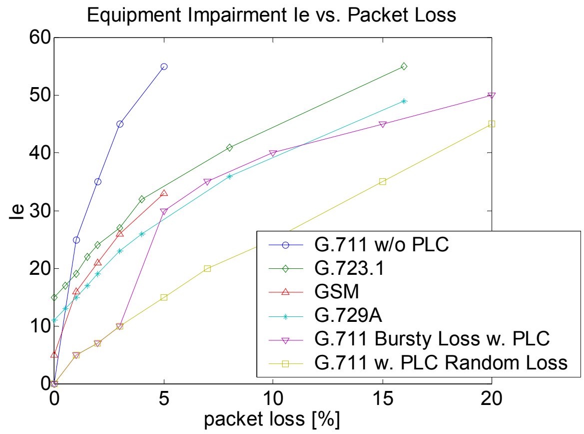


Fig. 6: Equipment impairment as a function of the packet loss [17]

Predicting transmission quality

If the mouth-to-ear delay, echo loss, encoding scheme, and packet loss are known, the transmission quality of a packetized voice call can be derived as shown in Figure 7.

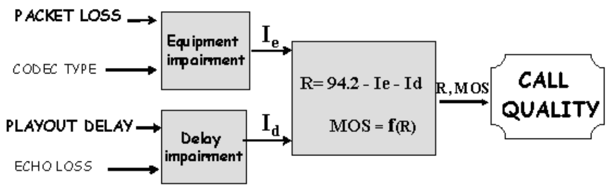


Fig. 7: Calculating voice transmission call quality using E-model methodology.

3.1.3 Assesing time varying of the call

The E-model does not take onto account the dynamics of a transmission but relies on static transmission parameters. A natural approach is to divide the call duration into fixed time intervals and assess the quality of each interval independently. This method for assessing time-varying quality of a call was proposed in [14]. There is one further important parameter that influences these calculations, namely the time interval for which the average playout delay and the average loss is calculated. It has been assumed that the time window of 10 seconds is sufficient because it is within the recommended length for PESQ algorithm [18].

The average playout delays and average packet loss is calculated by the playout buffer module for every 10 seconds of a transmission as shown in Figure 8. The corresponding delay impairments (assuming given echo loss), equipment impairments (assuming given codec type), and the resulting rating factor R are shown in Figure 8.

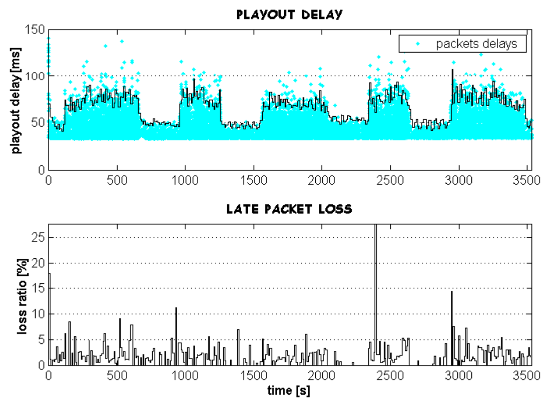


Figure 8: Average playout delays and packet loss for each 10 seconds of a call.

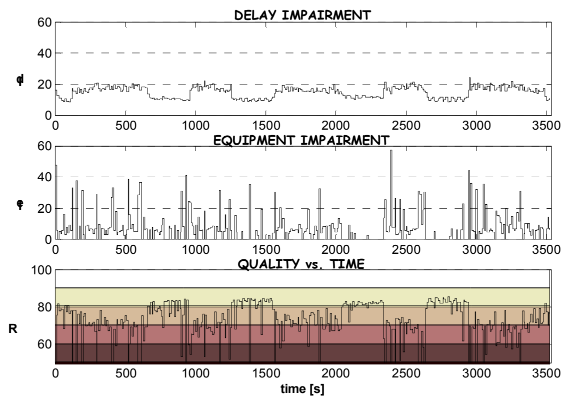


Figure 9: Corresponding transmission impairments and time varying quality of a call (rating R).

3.2 Assessing user satisfaction

Using the formula in equation (2), we created contours of quality as a function of delay and loss. Such quality contours determine the rating factor R for all possible combinations of loss and delay, with their shape being determined by both impairments I_d and I_e. They give a measure of the impact of packet loss and compression scheme on speech quality and the effect of delay and echo on interactive conversations.

Figure 10 shows those quality planes for G.711 encoding scheme (assuming bursty loss of packets) and for five different echo loss levels (TELR=45, 50, 55, 60, 65).

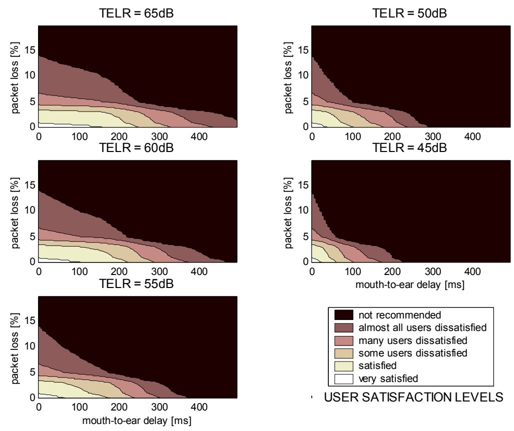


Figure 10: User perceived quality planes for G.711 encoding w. PLC (bursty loss).

Such quality planes can be used to assess overall user satisfaction as follows:

The playout buffer module calculates playout delays and resulting packet loss with the use of a specific playout algorithm. Those playout delays and packet losses can be mapped on loss/delay plane with quality contours on it as shown in Figure 11. Each dot on Figure 10 corresponds to average playout delay and average late packet loss for 10 seconds of the transmissions. This mapping is directly related to user perceived quality as shown on Figure 12.

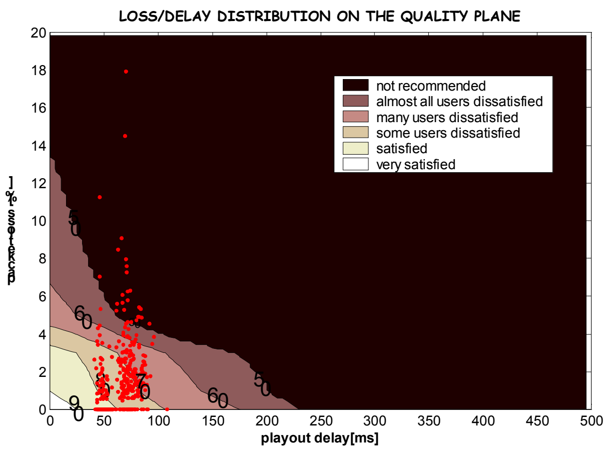


Fig. 11: Distribution of playout delays and packets loss on the quality plane (codec G.711 w. PLC, bursty loss, echo level TELR = 45dB).

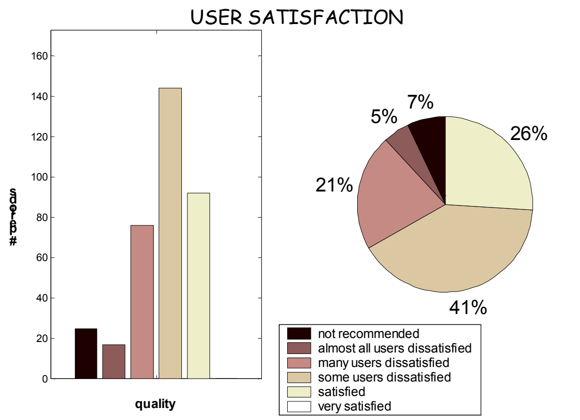


Fig. 12: User satisfaction (codec G.711 w. PLC, bursty loss, TELR = 45dB).

As shown in Figure 11, using the specific algorithm, with the specific codec and the specific echo loss:

- an average user would be satisfied 26% of the time
- some users could be dissatisfied 41% of the time
- many users would be dissatisfied 21% of the time
- almost all users would be dissatisfied 5% of the time
- during 7% of the time quality was not acceptable at all.

4 Evaluation various buffer algorithms.

4.1 Experimental setup.

An one-way VoIP session was established between two wireless hosts (VoIP SENDER and VoIP RECEIVER), via the Access Point (AP), in an 802.11b WLAN (Fig.13).

A number of wireless stations were used to generate background UDP traffic. This was accomplished using the Iperf traffic generator [19]. The stations generated UDP packets of length 1024 bytes at a transmission rate of 50 fps.

Voice traffic was generated using RTPtools [20].

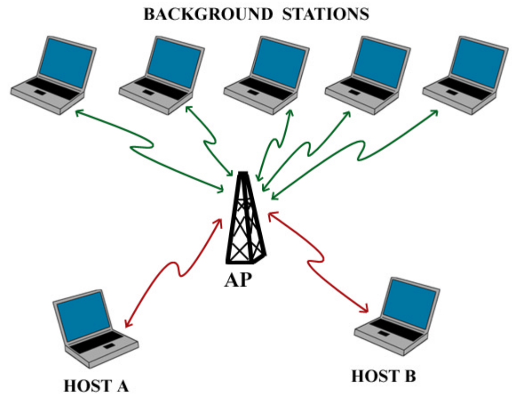


Fig. 13: Measurement setup

The VoIP sender sent voice packets of 80 bytes every 10 ms (i.e. G.711 codec) during voice activity. No packets were generated during silence periods. A sequence of alternating active and passive periods was used following the ITU-T P.59 recommendation [21] with an exponential distribution of talkspurts and gaps (with mean values of 1004ms and 1587ms respectively). The duration of the test was one hour during which time all experimental data (packet arrival times, timestamps, sequence numbers, and marker bits) were collected at the receiving terminal and processed later (off-line) with a program that simulated the behaviour of various playout algorithms. Since the terminal clocks were not synchronized, the clock skew was removed using Paxon's algorithm [22].

The influence of the background traffic on the delay and delay variation is shown in Figure 14.

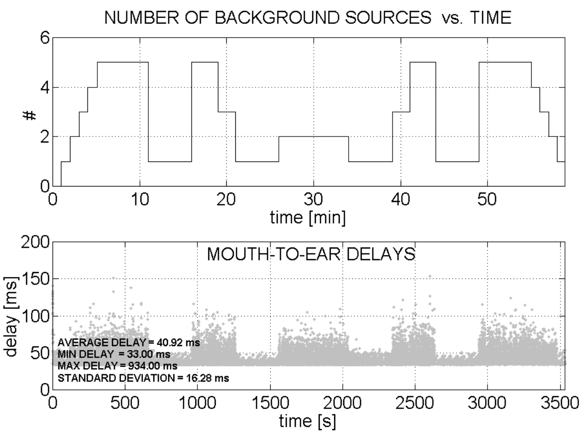


Fig. 14: Influence of the background traffic on delay and jitter.

4.2 Effect of various buffering schemes on loss/delay trade-off.

It is possible to distinguish four groups of playout buffer algorithms:

- reactive algorithms that perform continuous estimation of network delays and jitter to calculate playout deadlines [2], [3]
- histogram-based algorithms that maintain a histogram of packet delays and choose the optimal playout delay from that histogram [4], [5]
- algorithms that monitor packet loss ratio or buffer occupancy and adjust the playout delay accordingly [23]
- algorithms that aim in maximizing user satisfaction [24]

We have proposed a new playout buffer algorithm [25] that extends the reactive approach. In our solution, the estimator gain is updated with each incoming packet according to the observed delay variations. When variations in network delays are high, which implies that network conditions are rapidly changing, the value of gain is set low, and vice-versa. We claim that with higher-quality estimates of network delays, our algorithm adapts quicker to changing network conditions, which reduces the frequency of late packets and the amount of buffering delay. We have tested this algorithm in the fixed Internet [26][27] comparing its performance with the performance of the basic Ramjee's algorithm [2]. In contrast to previous work, in this paper we evaluate this algorithm in an IEEE802.11b WLAN environment comparing its performance with another reactive algorithm (Bolot's [3]) and two histogram-based algorithms ([4],[5]).

In reactive algorithms (Ramjee and Bolot), it is the β parameter (ranging from 2 to 4) that controls the loss/delay trade-off. In histogram based-algorithms (Moon and Concord), we can control it by specifying the desired packet loss rate (in the range from 0% to 10%).

The figures below show the trade-off between average buffering time and average late packet loss rate for various adaptive playout schemes.

In figure 15, the solid lines represent the performance of Ramjee's basic algorithm with fixed α (0.8, 0.9 and 0.998002), the line with triangles represents the performance of Bolot's algorithm and the line with circles represents the performance of our new algorithm.

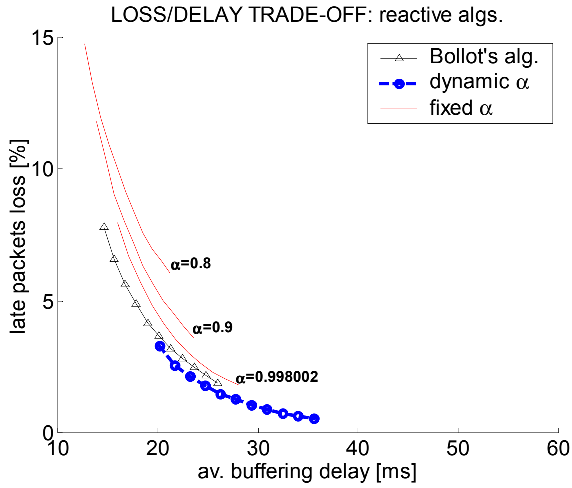


Fig. 15: Late packet loss rate vs. average buffering delay for reactive algorithms

As can be seen, our new algorithm achieves a better loss/delay trade-off than reactive algorithms (α=0.8, 0.9, 0.998002 or specified by Bolot's equations), for the full range of β values.

In figure 16, the solid lines represent the performance of Moon's algorithm with a fixed histogram window, the line with triangles represents the performance of the Concord algorithm, and the line with circles represent the performance of our algorithm with dynamic α for comparison purposes. Again, our new algorithm achieves a better loss/delay trade-off than the histogram-based algorithms (number of samples in the histogram = 100, 200, 400, 1000 or the whole trace in the "Concord" case).

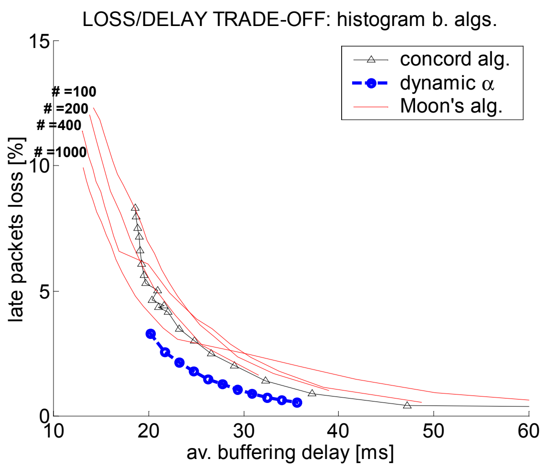


Fig. 16: Late packet loss rate vs. average buffering delay for histogram-based algorithms.

4.4 Effect of various buffering schemes on subjective quality.

Based on the E-model methodology described in section 3, we assessed time varying quality of the call and subjective user satisfaction. First we calculated the average playout delay and average packet loss for 10 second periods of the transmission. Assuming G.711 encoding with PLC, random loss, and echo cancellation implemented (TELR=65dB) we calculated delay impairments and equipment impairments and finally found time factor R.

The figures below show average playout delays, average packet loss and corresponding rating factor R for different algorithms. Overall user satisfaction over a one-hour period was obtained from delay/loss distribution on the user perception quality plane.

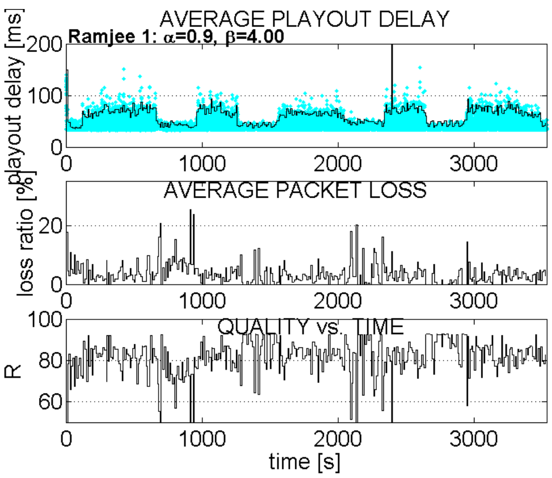


Fig. 17: Time varying playout delay, packet loss and quality of the call with the Ramjee's alg. (α = 0.9)

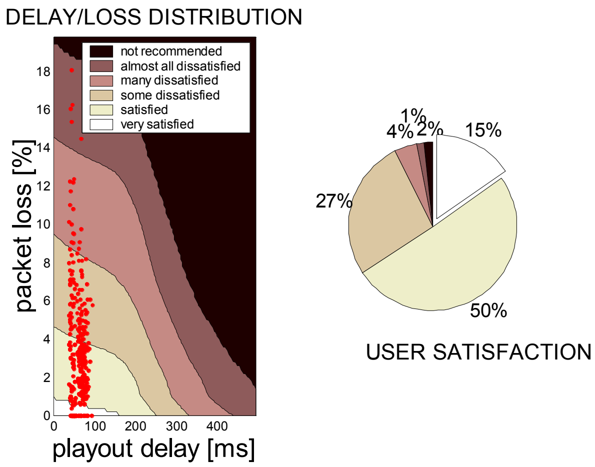


Fig. 18: Distribution of playout delays and packet loss on the quality plane with the Ramjee's alg. (α = 0.9) *and resulting user satisfaction*

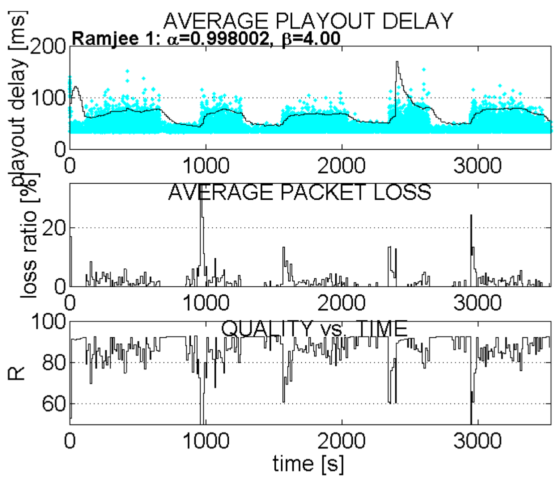


Fig. 19: Time varying playout delay, packet loss and quality of the call with the Ramjee's alg. (α = 0.998002)

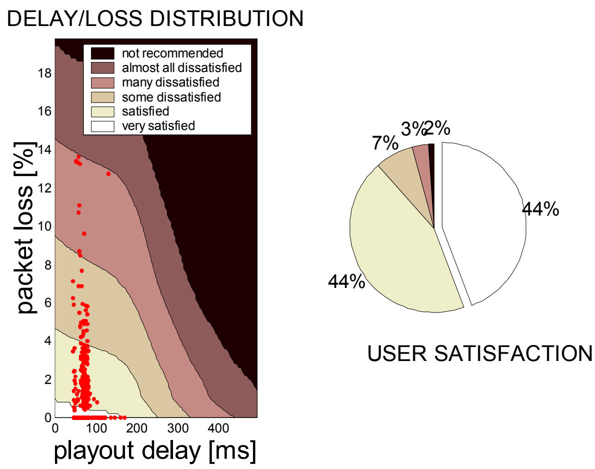


Fig. 20: Distribution of playout delays and packet loss on the quality plane with the Ramjee's alg. (α = 0.998002) and resulting user satisfaction

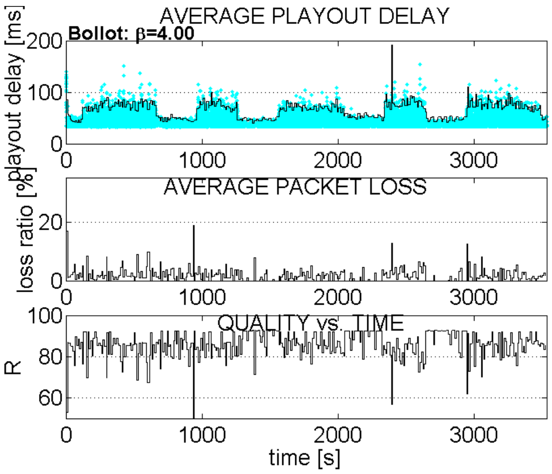


Fig. 21: Time varying playout delay, packet loss and quality of the call with the Bolot's alg.

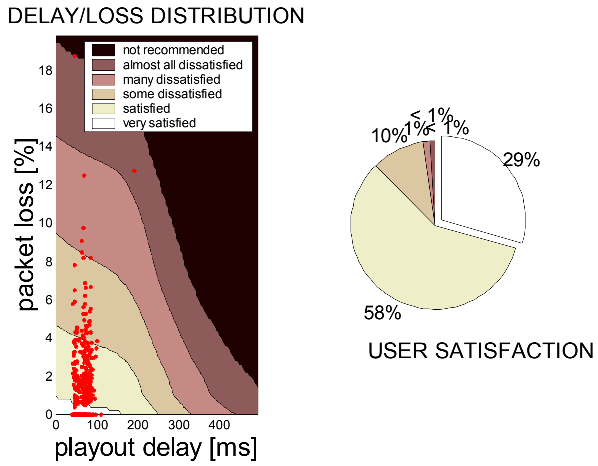


Fig. 22: Distribution of playout delays and packet loss on the quality plane with the Bolot's alg. and resulting user satisfaction

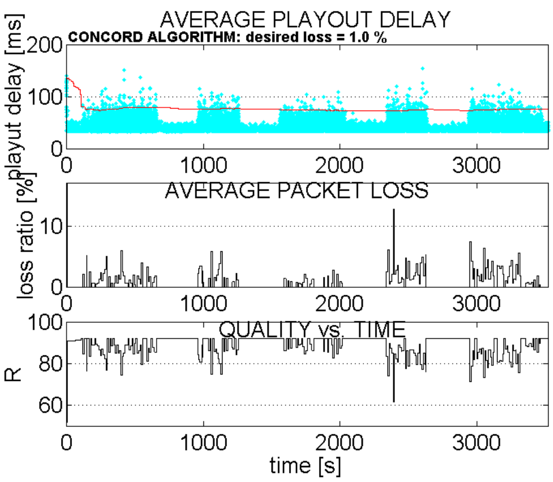


Fig. 23: Time varying playout delay, packet loss and quality of the call with the "Concord" alg.

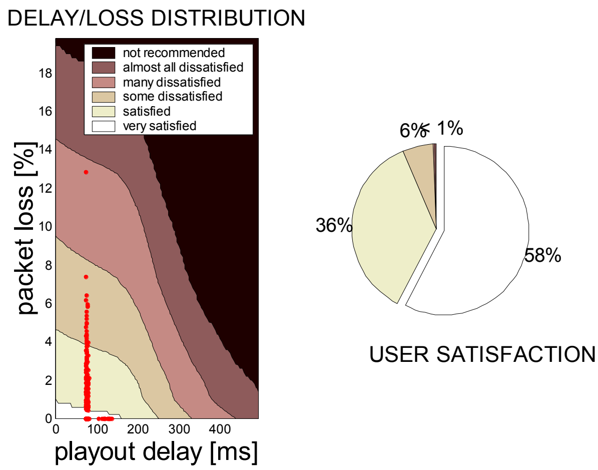


Fig. 24: Distribution of playout delays and packet loss on the quality plane with the "Concord" alg. and resulting user satisfaction

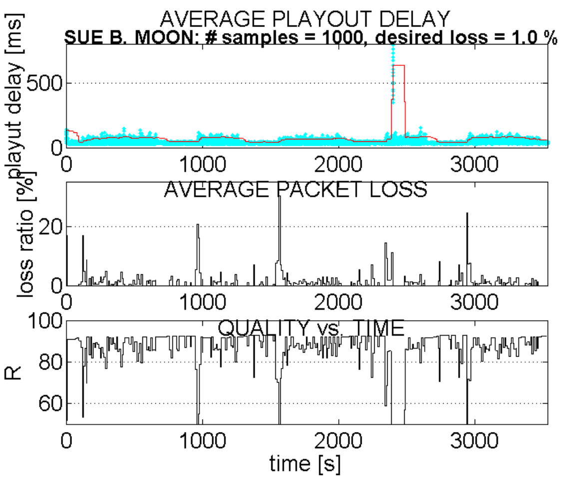


Fig. 25: Time varying playout delay, packet loss and quality of the call with the Moon's alg.

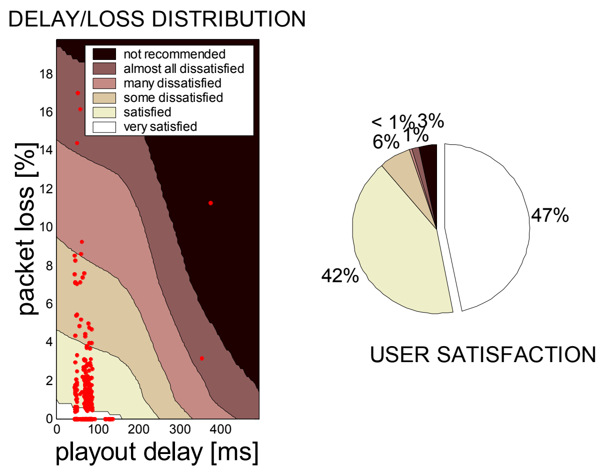
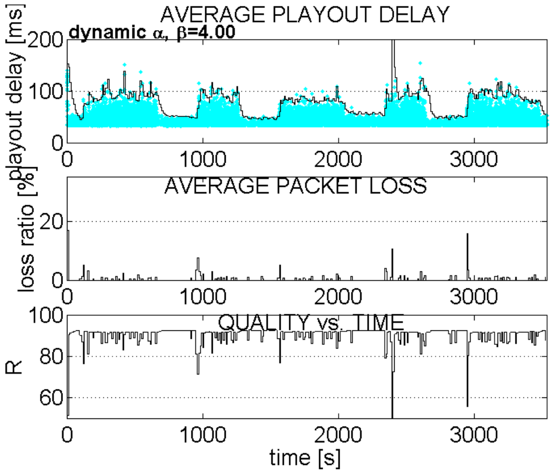


Fig. 26: Distribution of playout delays and packet loss on the quality plane with the Moon's alg. and resulting user satisfaction



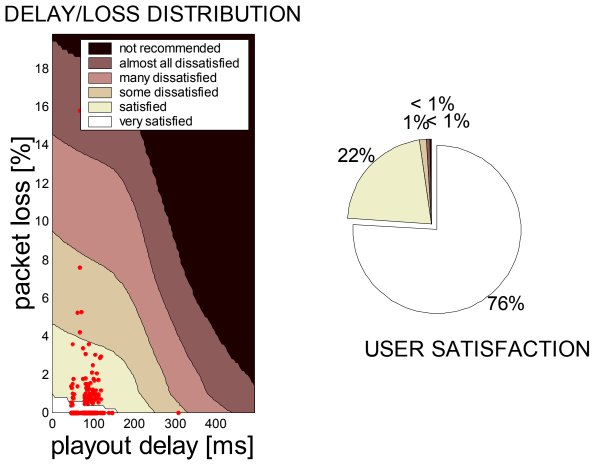


Figure 27: Time varying playout delay, packet loss and quality of the call with the "dynamic α" alg.

Figure 28: Distribution of playout delays and packet loss on the quality plane with the "dynamic α" alg. and resulting user satifaction

These results above show that the new adaptive buffering scheme with dynamic α gave very good user satisfaction 76% of the time, compared to the basic algorithm with fixed α at 44% (α=0.998002), Bolot 29%, Moon 47% and Concord 58%. This indicates that the dynamic α approach responds well to the fast variations that are expected in a WLAN environment.

5. Conclusions

Delay, echo, encoding scheme and packet loss all influence perceived quality of conversational speech. Therefore, the choice of a buffer algorithm cannot be solely based on statistical loss/delay trade-off metrics. Also subjective "listening tests" or the new ITU-T PESQ method, which don't consider the effect of mouth-to-ear delay are inappropriate. We proposed a method for assessing VoIP call quality by extending the ITU-T E-model concept. This method provides a direct link to perceived conversational speech quality by estimating user satisfaction from the combined effect of information loss, delay and echo. We compared various adaptive buffering algorithms taking into account: average buffering delay, late packet loss ratio and user perceived quality as measured by the proposed method. We observed that histogram-based algorithms are not capable of very rapidly increasing the buffering delay during congestion and quickly reducing it when congestion has passed. Also reactive algorithms (that rely on fixed estimator gain) tend to either react too quickly to transient noise conditions (when the estimator gain is small) or ignore persistent changes in performance (when the estimator gain is high).

The adaptive playout algorithm with dynamic estimator gain predicts and follows network delays more effectively in the wireless LAN environment than existing reactive and histogram based algorithms. Results show that with the dynamic estimator gain, one can achieve better delay/loss trade-off, better call quality, and better user satisfaction.

References

[1] ITU-T Recommendation G.1020, "Performance parameter definitions for quality of speech and other voiceband applications utilising IP networks", November '03

[2] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide-area networks", in Proceedings of the IEEE Infocom, June '94

[3] J.C. Bolot and A. Vega-Garcia, "Control mechanisms for packet audio in the Internet". In Proceedings of the IEEE Infocom '96, April '96

[4] N. Shivakumar, C. J. Sreenan, B. Narendran, and P. Agrawal, "The Concord algorithm for synchronization of networked multimedia streams". In Proceedings of the IEEE International Conference on Multimedia Computing and Systems, May '95

[5] S. B. Moon, J. Kurose, and D. Towsley, “Packet audio playout delay adjustment: Performance bounds and algorithms”, ACM/Springer Multimedia Systems, Vol. 6, January ‘98

[6] Y. J. Liang, N. Färber, and B. Girod, "Adaptive Playout Scheduling and Loss Concealment for Voice Communication over IP Networks", IEEE Transactions on Multimedia, Vol. 5, Issue 4, December ‘03

[7] ITU-T Recommendation P.800, “Methods for subjective determination of transmission quality”, August, ‘96

[8] ITU-T Recommendation P.861, “Objective quality measurement of telephone-band (300-3400 Hz) speech codecs”, February, ‘98 (withdrawn)

[9] ITU-T Recommendation P.862, “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs”, February, ‘01

[10] N. O. Johannesson: “The ETSI Computation Model: A Tool for Transmission Planning of Telephone Networks”, IEEE Communications Magazine, pp. 70-79, January ‘97

[11] ITU-T Recommendation G.107, “The E-Model, a computational model for use in transmission planning”, March, ‘03

[12] Telecommunications Industry Association, “Voice Quality Recommendations for IP Telephony – TIA/EIA/TSB116”, ‘01

[13] ITU-T Recommendation G.109, “Definition of categories of speech transmission quality”, September ‘99

[14] A. P. Markopouolu, F. A. Tobagi, M. J. Karam, “Assessment of VoIP Quality over Internet Backbones”, Proceedings of the IEEE Infocom, June, ‘02

[15] ITU-T Recommendation G.114, “One-way transmission time”, May, ‘03

[16] ITU-T Recommendation G.131, “Talker echo and its control”, November ‘03

[17] ITU-T Recommendation G.113, “Transmission impairments due to speech processing Appendix 1”, May, ‘02

[18] L. Sun, E. C. Ifeachor, “Prediction of Perceived Conversational Speech Quality and Effects of Playout Buffer Algorithms”, Proceedings of the IEEE ICC, May, ‘03

[19] Iperf: http://dast.nlanr.net/Projects/Iperf/

[20]RTPtools: http://www.cs.columbia.edu/IRT/software/rtptools/

[21]ITU-T Recommendation P.59, “Artificial conversational speech”, March ‘93

[22] Sue. B. Moon, P. Skelly, and D. Towsley, "Estimation and removal of clock skew from network delay measurements," in Proc. of IEEE INFOCOM'99, March ‘99

[23] D. L. Stone, K. Jeffay, “An empirical study of delay jitter management policies”, ACM/Springer Multimedia Systems Journal, Vol. 2, No. 6, January, ‘95

[24] A. P. Markopoulou, PhD Dissertation: “Assessing the Quality of Multimedia Communications over Internet Backbone Networks”, Stanford University, 2003

[25] M. Narbutt, L. Murphy, “VoIP Playout Buffer Adjustment using Adaptive Estimation of Network Delays”, Proceedings of the 18-th International Teletraffic Congress – ITC-18, p.1171-1180, September ‘03

[26] M. Narbutt, L. Murphy, “A New VoIP Adaptive Playout Algorithm”, Proceedings of the IEE QoS 2004 Conference, p.99-103, March ‘04

[27] M. Narbutt, L. Murphy, “Improving Voice over IP Subjective Call Quality”, IEEE Communications Letters, Vol. 8, Issue 5, p. 302-304, May, ‘04