User-Perceived Quality Assessment for Multimedia Applications

M. Ivanovici* and R. Beuran†
*University "POLITEHNICA" Bucureşti, România, E-mail: mivanovici@alpha.imag.pub.ro
†Japan Advanced Institute of Science and Technology, Japan, E-mail: razvan@jaist.ac.jp

***Abstract*— Multimedia applications are more and more widely used today. Internet will become the dominant distribution media for radio and television. In order to offer the same quality provided by classic broadcasting equipment, the real-time requirements of multimedia applications must be met by the network. Therefore, the need to accurately measure their performance and its dependence on network conditions emerged with urgency.**

We focused our study on two multimedia applications: voice over IP (or Internet telephony) and video streaming. The user-perceived quality of the voice or video signals is influenced by the amount of quality degradation at network level. We experimentally determined this dependency by using a network emulator and a monitoring system that we designed and implemented.

I. Introduction

Multimedia applications are the class of network applications with real-time requirements that is currently the most widely deployed over the Internet. In addition, personal computer and consumer technology analysts consider that the Web is to become an increasingly dominant distribution method for movies and television (see, for example, [3]). The non-negligible requirements of such applications in terms of network conditions lead to a wider recognition of the issues related to Quality of Service (QoS). We consider QoS to be the fidelity of a system's observable behaviour to expectations: one can only assess quality by comparing the result of a measurement with the expected value for that measurement. Determining the performance characteristics of a network system is the first step in understanding the application-level behaviour. This must be followed by an evaluation of the user-perceived quality (UPQ) for that particular application, and the establishment of the relationship with the measured QoS parameters.

All network applications require a minimum QoS level in order to run according to user expectations [4], [5]. Network elements along the path cause degradation that accumulates. There is a maximum end-to-end quality degradation (denoted by ΔQ) within which the network must deliver the application traffic for it to run in a satisfactory manner. The number of systems designed to correlate the quality differentiation provisioned by networks with the UPQ for specific applications is reduced. Knowing the requirements of multimedia applications such as Voice over IP (VoIP) or video streaming allows predicting whether a certain connection is valid for this particular type of real-time applications, and what will be the perceived quality for that application.

A key issue in the context of UPQ assessment is the understanding of the fact that network environments perturb application behavior by delaying and dropping application traffic. Networks are therefore degraded environments, and quality degradation in the network is reflected in the quality of the voice or video signal, as it is perceived by the user.

There are three steps to take in order to assess application UPQ: (i) observe the application behavior at the end-node level, (ii) accurately measure the quality degradation experienced by the application traffic and (iii) correlate the above. A general setup is depicted in Figure 1.

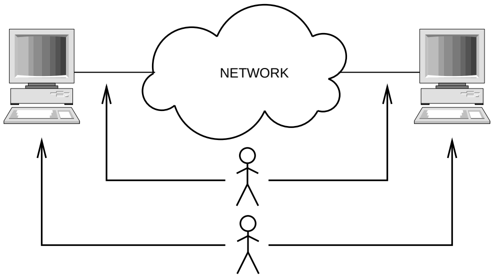

Figure 1. Observing end-to-end application performance, and measuring the quality degradation in the network.

Scientific method requires the use of objective metrics to perform both the network and application-level performance assessments. In case of network quality degradation there is already a series of widely-used metrics [12], [19]: one-way delay [1], one-way packet loss [2] and throughput. However, when application performance must be determined, each application class requires the definition of specific metrics that take into account the nature of that application. For example, for VoIP one can use the Perceptual Evaluation of Speech Quality

(PESQ) score [14]. However there is no standardization to date concerning video applications. The next section will present the current status in this area.

A. State of the art for video quality assessment

For the quantification of the quality of a video sequence, there are two types of metrics: subjective and objective. Subjective video quality measurements are time consuming and must meet complex requirements (see the ITU-T recommendations [7], [15], [8], [13]) regarding the conditions of the experiments, such as viewing distance and room lighting.

On the other hand the objective metrics can be implemented as algorithms and are human-error free. They are either based on models of the human-vision system or on distance measures, such as the Root Mean Square Error (RMSE) or the Peak Signal-to-Noise Ratio (PSNR). However these simple measures do not capture the user-perceived degradation in the video signal. Image attributes like sharpness and colorfulness should be taken into account [26], [25] for that purpose.

The UPQ metrics can also be with reference, when the sequence at receiver is compared to the original sequence at transmitter, or without reference, when only the sequence at receiver is analyzed.

The Video Quality Experts Group (VQEG) reported on the perceptual video quality measurement algorithms [23]. A survey of video-quality metrics based on models of the human vision system can be found in [6]. Several no-reference blockiness metrics are studied and compared in [27]. Most of the existing metrics for the video quality quantify the degradation introduced by the compression algorithm itself or due to the frame rate that is used. There are no metrics or studies that objectively assess the degradation in video quality caused by the packet loss at network level. Therefore we proposed a set of UPQ metrics for video applications that take into account particularly this aspect.

II. UPQ METRICS FOR VIDEO APPLICATIONS

The metrics we used for the assessment of the performance of video-streaming applications are described in detail in [17]. The two objective metrics are reference-based, i.e. the original and the degraded video sequences are compared in order to assess the perceptual degradation that occurred.

The number of altered video frames (NAF) indicates how many frames—from the ones received and rendered—are affected by impairments. The number of dropped video frames (NDF) represents the difference between the number of frames in the original video sequence at transmitter (server) and the number of video frames that are effectively rendered at the receiver end (client). This number indicates how many frames are missing at receiver when the MPEG video stream was incomplete because of packet loss in the network.

Note that these metrics are not independent. A larger packet loss in the network may lead to a smaller number of altered frames being received, since some of them are lost completely; a smaller NAF in this case doesn't of course mean an improvement in quality. To cope with this phenomenon we summed the aforementioned metrics as the total number of affected video frames (TNAF), i.e. the frames that are dropped or altered.

III. EXPERIMENTAL RESULTS

The multimedia applications we focused on are VoIP and video streaming. This section presents the experimental setup and representative results for each of the applications.

A. Test setup

The setup we used to perform these experiments is generic, i.e. it can be used to assess the performance of any network application. The detailed description of our monitoring system can be found in [4]. We remind here the basic facts.

We mirror the traffic on the link between two PCs that run the network application under study using FastEthernet taps. This traffic is fed into programmable Alteon UTP network cards (NICs). From each packet the information required for the computation of the network QoS parameters is extracted and stored in the local memory as packet descriptors. The host PCs, which control the programmable NICs, periodically collect this information and store it in descriptor files. This data is then used to compute off-line the following network QoS parameters: one-way delay and jitter, packet loss and throughput. We can calculate instantaneous or average values, and various histograms.

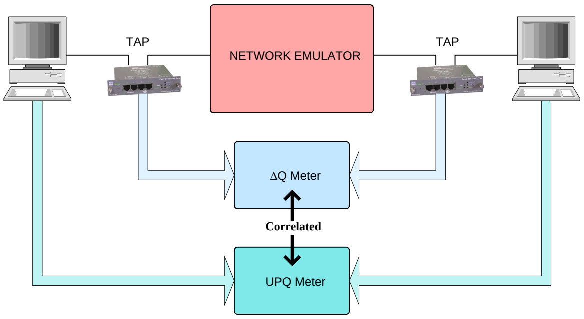

Figure 2. Experimental setup.

Our setup makes use of a network emulator (the NIST Net network emulator [24]) to study a wide range of controllable network conditions. Taps are used to monitor network traffic, which is fed into the block "ΔQ Meter" so as to accurately measure the quality degradation in the network. The block “UPQ Meter” is application-dependent and makes it possible to assess the UPQ for the particular application under study.

We modified the source code of both multimedia applications that we studied in order to enable the clients to save the audio and the video sequence received. This allowed us to quantify the degradation that occurred at user level in an objective manner.

B. *VoIP*

VoIP is a widely-used interactive network application. The bandwidth requirements of speech transmission are low (64 kb/s voice data maximum), but interactivity implies high sensitivity to delay and jitter. We haven’t studied the influence of one-way delay on VoIP UPQ because these requirements are generally known [16], [21]: a mouth-to-ear delay of up to 150 ms gives good interactivity, a delay between 150 and 400 ms is acceptable, and delays higher than 400 ms are unacceptable. Therefore we performed only uni-directional tests, which focus on the perceived quality of the speech itself depending on packet loss and jitter. In each case the UPQ was determined by means of the PESQ score [14]; for its computation we used an implementation supplied by Malden Electronics Ltd. [18].

In our tests with VoIP we used a freeware application, namely Speak Freely version 7.6a [24]. The application doesn’t do any of the following: silence suppression, reordering of out-of-order packets, packet loss concealment. We used a de-jittering buffer of 80 ms. We present here a study of the region with loss rates between 0 and 15% and average jitter values ranging from 0 to 75 ms, since quality becomes unacceptable within these boundaries already. Five series of tests were run for each codec in order to collect the data used for the results shown below. A detailed description of the test conditions is available in the technical report [5].

The four codecs we performed experiments with are: G.711, G.726, GSM and G.729. The G.711 codec [9] sends data at 8 kHz with 8 bits per sample, resulting in a data rate of 64 kb/s. The sound is in PCM format, encoded using the μ-law. The G.726 codec [10] converts a 64 kb/s μ-law or A-law PCM channel to and from 40, 32, 24 or 16 kb/s channels. In our application only the 32 kb/s encoding is available. The GSM (Global System for Mobile telecommunications) codec [20] uses linear predictive coding (LPC) to compress speech data at 13 kb/s. The G.729 codec [11] is frequently used for VoIP communication. It sends data at 8 kb/s using conjugate-structure algebraic-code-excited linear-prediction (CSACELP).

The results on G.711 were previously presented in [5], but further experiments allowed us to realize a comparison with the low bit-rate codecs that we studied subsequently; this comparison is presented next.

The basic characteristics of the codecs we used are summarized in Table I: transmission rates (VoIP data rate, network rate and packet rate), as well as network packet sizes, when using RTP as a transport protocol are provided.

Codec	Data rate [kb/s]	Packet size [bytes]	Network rate [kb/s]	Packet rate [packets/s]
G.711	64	378	75.6	25
G.726	32	382	38.2	12.5
GSM	13	190	19	12.5
G.729	8	170	17	12.5

TABLE I
CODEC CHARACTERISTICS.

According to [22] the relationship between PESQ scores and audio quality is the following: (i) PESQ scores between 3 and 4.5 mean acceptable perceived quality, with 3.8 being the PSTN[1] threshold—this will be termed as *good quality*; (ii) values between 2 and 3 indicate that effort is required for understanding the meaning of the voice signal—this will be named *low quality*; (iii) scores less than 2 signify that the degradation rendered the communication impossible, therefore the quality is *unacceptable*.

Based on this information the figures 3 to 6 show for each codec the boundaries on QoS parameters that must be enforced in order to attain a certain quality level. For example, G.711 provides good quality as long as loss rate is below 4% and average jitter doesn’t exceed 30 ms. The same codec will provide low but acceptable quality if loss rates are roughly between 4 and 14% and jitter is between 30 and 45 ms. Outside these bounds the quality will be unacceptable. Note that G.711 is the only codec amongst those tested that also provides very good (PSTN) quality.

A general conclusion is that the codec G.711 performs better than the other codecs as long as network conditions are good (loss rate smaller than 3% and jitter below 20 ms). The codec G.726 gives better results than the GSM codec in the entire range of loss rates and jitter that we have studied, at the expense of a transmission rate that is 2.5 times larger. However their behaviour under the influences of loss and jitter are similar. G.729 seems to be the most robust codec in the range of network conditions under study. It provides almost the same perceived quality (always within the bounds of “low” quality) throughout almost 90% of the loss-jitter space. A decrease only slightly larger than 1.5 is observed from zero loss, zero jitter conditions to a loss of 15% and a jitter of 75 ms, which is considerably better than what the other codecs offer.

[1] Public Switched Telephone Network.

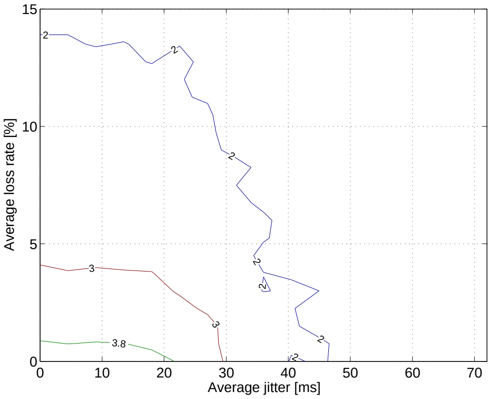

Figure 3. Contour plot of the boundaries between quality levels for the G.711 codec.

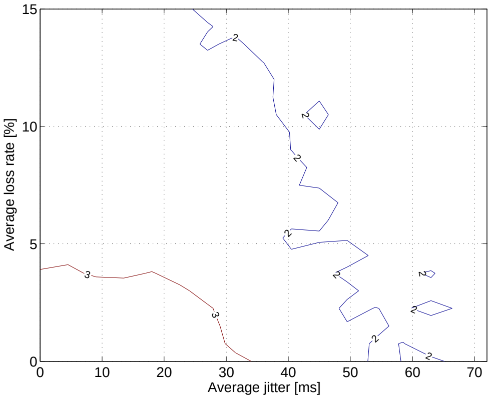

Figure 4. Contour plots of the boundaries between quality levels for the G.726 codec.

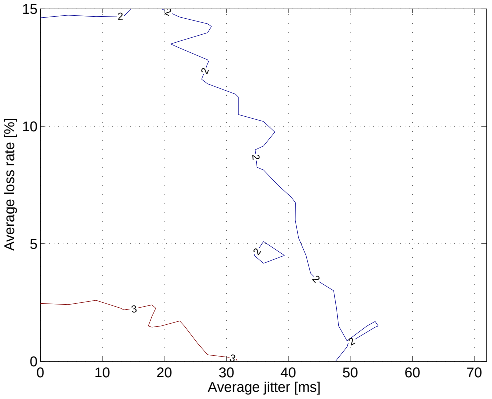

Figure 5. Contour plot of the boundaries between quality levels for the GSM codec.

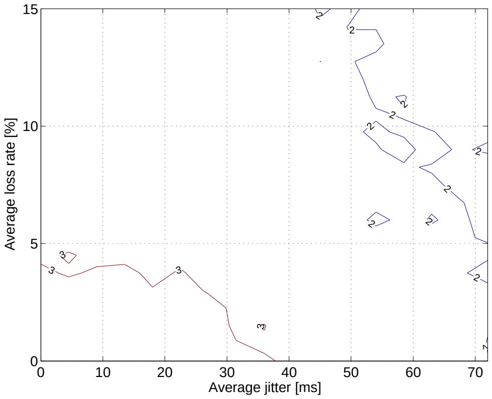

Figure 6. Contour plot of the boundaries between quality levels for the G.729 codec.

Tables II and III show a codec comparison from the point of view of PESQ score thresholds. We classify here the codecs based on the coverage of the area corresponding to a certain quality level with respect to the area of the studied loss-jitter space. In what follows we present two such classifications, one for the area of *at least good quality* (PESQ scores larger or equal to 3) and one for the area of *at least low quality* (PESQ scores larger or equal to 2). In these tables codec performance diminishes from the top to the bottom row. Note however that this classification doesn't take into account the bit-rates of each codec, which are also important when making the trade-off between perceived quality and network utilization efficiency.

For demanding users that require at least good quality of the speech signal, one can choose the codec based on table II. Less demanding users, for which low quality is sufficient, can use VoIP in a wider range of network conditions, by choosing the appropriate codec from table III.

Codec	Good quality coverage
G.729	10.48%
G.726	9.62%
G.711	9.00%
GSM	5.06%

TABLE II

CODEC CLASSIFICATION BASED ON GOOD QUALITY COVERAGE.

Note that the codec G.729 is on the first position in both tables, meaning that it performs best in our study. Given that it is also the codec with the lowest bit-rate, we consider it as the codec of choice in almost any situation.

Codec	Low quality coverage
G.729	88.16%
G.726	60.33%
GSM	51.25%
G.711	41.7%

TABLE III

CODEC CLASSIFICATION BASED ON LOW QUALITY COVERAGE.

C. *Video streaming*

For video streaming tests the only network parameter that was varied was packet loss. Using the network emulator packet loss was introduced only in the server-client direction, on the direction of the video data flow. Packet loss values ranged from 0 to 1%.

We used two standard MPEG-4 video sequences for testing: “football” and “train”. These video sequences are 10 second long, with 250 frames of 320x240 pixels, resulting in an average transmission rate of approximately 1 Mb/s.

An example of altered MPEG-4 video frame from the “football” video sequence is presented in Figure 7. Note that the degradation of the video frame occurs mainly in the regions involving moving objects.

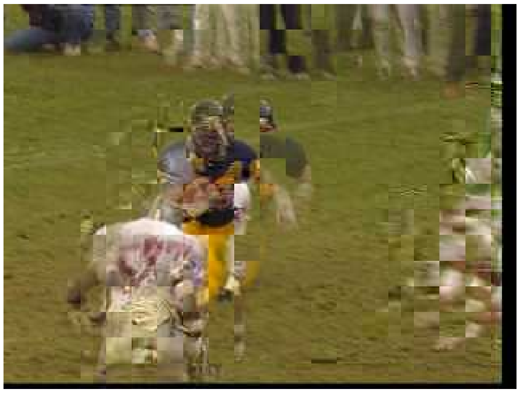

Figure 7. Degraded video frame from the “football” video sequence.

Figure 8 shows the percentage of altered frames, and Figure 9 the percentage of dropped frames as function of packet loss for the two video sequences.

The decreased number of altered frames when the loss rate is larger than 0.8% (see Figure 8, the “train” sequence) is the consequence of a larger number of dropped video frames. This is an indication of the fact that some frames are completely lost or that they are so severely degraded frames that the system can no longer render them.

To prevent a misleading reading of the figures above, we used the third metric, TNAF, and plotted the total number of affected video frames as a function of packet loss at network level. The monotonic increase of TNAF can be observed in Figure 10. In this figure one can observe, for example, that dropping 0.6% of the packets at network level affects 60% of the video frames. This shows how vulnerable to packet loss an MPEG-4 video streaming application is. Note that video frames in MPEG-4 encoding are of three categories, and the effects of losing a frame from a category is different, as follows. Losing one packet containing the information of an I (intra) frame from the MPEG-4 stream implies the degradation of all the following P (predictive) or B (bi-directional predictive) frames. Altering the information of a P frame implies only the degradation of another adjacent frame. Altered B frames do not cause the degradation of other video frames.

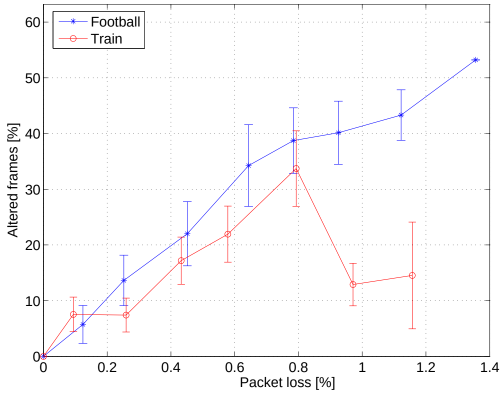

Figure 8. Number of altered frames as function of packet loss.

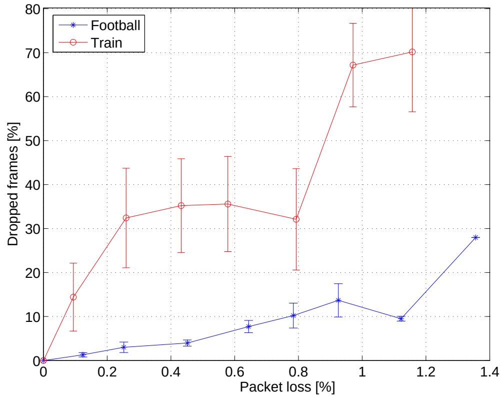

Figure 9. Number of dropped frames as function of packet loss.

IV. CONCLUSIONS

The work we carried out quantifies the dependency that exists between network conditions and perceptual quality network communication applications in an objective manner. In this paper we focused on two multimedia

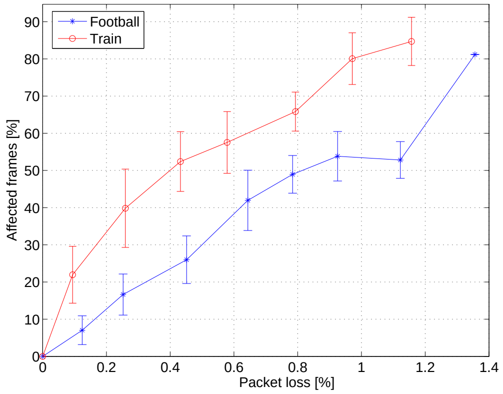

Figure 10. Number of affected (altered or dropped) frames as function of packet loss.

applications: VoIP and MPEG-4 video streaming. We used the ITU-T PESQ score for VoIP UPQ and two objective reference-based metrics for the user-perceived quality of video sequences.

According to our study the codec G.729 appears to perform the best, providing good or acceptable quality for a wide range of network conditions, at the lowest bit rate. For MPEG-4 video streaming applications, the packet loss has a strong influence on the quality of the video signal. Loss percentages larger than 1% lead to unacceptable quality of the received video signal, or even to application failure.

The system we designed and implemented makes it possible to accurately measure network quality degradation, and objectively assess application UPQ in parallel. This permits to experimentally determine the dependence of UPQ on network quality degradation for any application.

V. Acknowledgments

We would like to acknowledge the valuable discussions concerning the study of VoIP that we had with Gerhard Heide, Chris Vowden and Annie Snelson. The support of Bob Dobinson and Neil Davies during our work provided essential guidelines. We also wish to thank our professor, Vasile Buzuloiu, for his prized feedback.

Brian Martin, Catalin Meirosu and Jamie Lokier had a significant contribution to the early development of the clock synchronization system and pioneered the reprogramming of the Alteon NICs, which are used in our QoS monitoring system. The authors would also like to thank Marius Malciu for the fruitful discussions and suggestions regarding the processing of video data.

References

[1] G. Almes, S. Kalidindi, M. Zekauskas, “A One-way Delay Metric for IPPM”, IETF RFC 2679, September, 1999.

[2] G. Almes, S. Kalidindi, M. Zekauskas, “A One-way Packet Loss Metric for IPPM”, IETF RFC 2680, September, 1999.

[3] Tim Bajarin, 2006 Predictions, Creative Strategies Insights, http://www.csinsights.com/index.php?action=pg_article&id=81

[4] R. Beuran, M. Ivanovici, B. Dobinson, N. Davies, P. Thompson, “Ntwork Quality of Service Measurement System for Application Requirements Evaluation”, International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS’03, Montreal, Canada, July 20-24, 2003, pp. 380-387.

[5] R. Beuran, M. Ivanovici, “User-Perceived Quality Assessment for VoIP Applications”, technical report, CERN-OPEN-2004-007, January 2004.

[6] C. J. van den Branden Lambrecht, “Survey of Image and Video Quality Metrics based on Vision Models” , presentation, August, 1997.

[7] ITU-R Recommendation BT.500, “Subjective quality assessment methods of television pictures” , ITU, 1998.

[8] ITU-R Recommendation J.140, “Subjective assessment of picture quality in digital cable television systems” , ITU, 1998.

[9] ITU-T Recommendation G.711, “Pulse Code Modulation (PCM) of voice frequencies”, ITU-T, 1993.

[10] ITU-T Recommendation G.726, “40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)”, ITU-T, 1990.

[11] ITU-T Recommendation G.729, “Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CSACELP)”, ITU-T, March 1996.

[12] ITU-T Recommendation I.380, “Internet Protocol (IP) Data Communication Service - IP Packet Transfer and Availability Performance Parameters”, ITU-T, February, 1999.

[13] ITU-T Recommendation J.143, “User requirements for objective perceptual video quality measurements in digital cable television” , ITU, 2000.

[14] ITU-T Recommendation P.862, “Perceptual Evaluation of Speech Quality (PESQ), An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Codecs”, ITU-T, February, 2001.

[15] ITU-T Recommendation P.910, “Subjective Video Quality Assessment Methods for Multimedia Applications” , ITU, 1996.

[16] ITU-T Recommendation Y.1541. 2001. “Network Performance Objectives for IP-Based Services”, ITU, draft, October.

[17] M. Ivanovici, “Objective Performance Evaluation for MPEG-4 Video Streaming Applications”, Scientific Bulletin of University ”POLTEHNICA” Bucharest, C Series (Electrical Engineering), January 2006.

[18] Malden Electronics Ltd., http://www.malden.co.uk.

[19] V. Paxson, G. Almes, J. Mahdavi, M. Mathis, “Framework for IP PerformanceMetrics”, IETF RFC 2330,May, 1998.

[20] M. Rahnema, “Overview of the GSM system and protocol architecture”, IEEE Communications Magazine, April 1993.

[21] Reijs, V. ”Perceived Quantitative Quality of Applications”, http://www.heanet.ie/Heanet/projects/nat_infrastruct/perceived.html.

[22] V. Servis, “Measuring speech quality over VoIP networks”, The TOLLY Group, December 2001.

[23] VQEG, “Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment”.

[24] B. C. Wiles, J. Walker, “Speak Freely”, http://www.speakfreely.org.

[25] S. Winkler, “Issues in Vision Modelling for Perceptual Video Quality Assessment” , Signal Processing, vol. 78, no. 2, pp. 231-252, October, 1999.

[26] S. Winkler, “Visual Fidelity and Perceived Quality: Towards Comprehensive Metrics” , Proc. SPIE Human Vision and Electronic Imaging, vol. 4299, pp. 114-125, San Jose, California, January, 2001.

[27] S. Winkler, A. Sharma, D. McNally, “Perceptual Video Quality and Blockiness Metrics for Multimedia Streaming Applications” , Proc. 4th International Symposium on Wireless Personal Multimedia Communications, pp. 553-556, Aalbord, Denmark, September, 2001.