Neural Network-based Interactive Multicriteria Decision Making in a Quality of Perception-oriented Management Scheme

G.D. Magoulas and G. Ghinea
Department of Information Systems and Computing
Brunel University
Uxbridge UB8 3PH,United Kingdom
Email: {George.Magoulas, George.Ghinea}@brunel.ac.uk

Abstract

In this paper we address the problem of bridging the application-network gap from a multicriteria decision making perspective. We have sought to use this approach to integrate results from our work on user-level Quality of Perception with the more technical characterization of Quality of Service. For this purpose, we apply the Multicriteria Decision-Making formalism by defining the criteria on the basis of which an appropriate tailored communication protocol is constructed, and we suggest a neural network-based scheme to solve the problem.

1 Introduction

The concept of Quality of Service (QoS) in distributed multimedia systems is indelibly associated with the provision of an acceptable level of application performance. In turn this performance is itself dependent on both the perceived, subjective quality of the application as well as its robustness to network congestion.

The focus of our research has been the enhancement of the traditional view of QoS with a user-level defined *Quality of Perception* (QoP). This is a measure, which encompasses not only a user's satisfaction with multimedia clips, but also his/her ability to perceive, synthesize and analyze the informational content of such presentations. As such, we have investigated the interaction between QoP and QoS and its implications from both a user perspective [7], as well as a networking angle [8].

Although the problem of multimedia application-level performance is closely linked to both the user perspective of the experience as well as to the service provided by the underlying network, it is rarely studied from an integrated viewpoint. Clearly, this is a very unsatisfactory state of affairs. In this paper, we try and rectify this situation by proposing a scheme for QoP management based on an integrated approach that takes into account both user-centric QoP and low-level QoS parameters. Such a scheme would then, by appropriate management of these QoS parameters, provide the potential of ensuring an optimum QoP in a distributed multimedia setting. But firstly, we shall take a closer look at the reasons behind the current segregated approach and some attempts that have been made to redress this imbalance.

The networking foundation on which current distributed multimedia applications are built either do not specify QoS parameters (also known as best effort service) or specify them in terms of traffic engineering parameters, such as delay, jitter, and loss or error rates. However, these parameters do not convey application-specific needs such as the influence of clip content and informational load on the user multimedia experience. As a result, the underlying network does not consider the sensitivity of the application performance to bandwidth allocation. There is, thus, an architectural gap between the provision of network-level QoS and application-level QoP requirements of the distributed multimedia applications. This gap causes distributed multimedia systems to inefficiently use network resources and results in poor end-to-end performance which in turn has a direct negative impact on the user experience of multimedia. The need for an integrated management scheme for QoP and QoS becomes therefore apparent.

Previous work in the area of subjective quality assessment of multimedia applications has concentrated on the entertainment side of multimedia. Accordingly, work has been done to appraise the enjoyment of multimedia presentations shown at differing frame rates [1], [6], while [20] and [21] examined the effect that random media losses have on user-perceived quality. The bounds within which lip synchronization can fluctuate without undue annoyance on the viewer's part have also been studied, [16], [17], as has the establishment of metrics for subjective assessment of teleconferencing applications [19].

In contrast to the relatively large body of work dealing with user perceived quality of multimedia presentations, there have been few instances in the literature of research being

0-7803-7044-9/01/$10.00 ©2001 IEEE

done in the area of bridging the application-network gap. This probably reflects the inherent difficulty of trying to link subjective sentiments about the quality of the presentation with the facts and figures of network parameters. In essence there are three approaches which can be identified:

1. The first approach tries to bridge the application-network gap *implicitly*. By this it is understood that there is no explicit mapping between application-level user requirements and the QoS provided by the network. What happens rather is that the user specifies, usually through a Graphical User Interface, his/her desired presentation quality [2], [13]. This is typically provided through the means of sliders or radio buttons via which the user would specify, for example, the desired playback frame rate, spatial resolution, as well as the acceptable synchronization delay between the audio and video streams.

2. In the second approach an *explicit* mapping linking application-level user requirements to network QoS is actually given. Such a mapping can either be defined on a per layer basis (such as a network to transport to session to application-layer mapping), or directly between application and network-level parameters [15].

3. The last approach is, in essence, a more restrictive version of the first. What happens here is that the user is played short-duration *probes* of differing deliverable qualities of the multimedia material in question and (s)he then specifies which of the given sample qualities is acceptable [11]. Apart from the obvious goal of polling user preferred multimedia quality the probe-based approach is advantageous from the point of view that it tests current network conditions. Thus, any choice that the user might make in as far as desired multimedia quality goes is guaranteed to be delivered - at least in the initial stages - by the network.

We have addressed the problem of bridging the application-network gap from a multicriteria decision making perspective. We have sought to use this approach to integrate results from our work on Quality of Perception with the more technical characterization of Quality of Service, with an ultimate aim of providing an adaptable communications protocol geared towards human requirements in the delivery of distributed multimedia. To achieve our goal, firstly, we exploit the *Dynamically Reconfigurable Stacks Project* (DRoPS) framework [4], which provides an infrastructure for the implementation and operation of multiple adaptable protocols. Secondly, we use a neural network approach to capture and represent the user preferences and then to suggest the most desirable alternative. Thirdly, we apply the Multicriteria Decision Making (MDM) formalism, [3], by defining the *criteria* on the basis of which an appropriate tailored communication protocol is constructed. Accordingly, the structure of this paper is as follows: Section 2 describes the framework that shall be used as a test bed for the new approach. Sections 3 and 4 show neural network-based management schemes for QoP and QoS parameters. Section 5 details the elements of an MDM method, called the Analytic Hierarchy Process (AHP), necessary for understanding our approach and describes the QoP-oriented ordering of QoS parameters. Lastly, conclusions are drawn in Section 6.

2 The DRoPS

Distributed guaranteed services need to incorporate capabilities for responding to QoP and QoS changes originating from the user/applications (as a result, perhaps, of the content of the application changing) or the system/network (as a result of changing network conditions), respectively. To achieve these changes, the networked multimedia system will require fast renegotiation protocols and adaptive mechanisms. The renegotiation protocols will rely on reliable and simple monitoring and recognition algorithms to detect requests for QoS changes or system degradations. The envisioned adaptive mechanisms should include update mechanisms for resource allocation in response to detection of system degradation. The challenge will be to make monitoring, detection and adaptation mechanisms efficient and fast [12].

In the *Dynamically Reconfigurable Stacks Project* (DRoPS) communication systems are composed of fundamental mechanisms, called *microprotocols*, which perform arbitrary protocol processing operations. The complexity of processing performed by a microprotocol is not defined by DRoPS and may range from a simple protocol function, such as a checksum, to a complex layer of a protocol stack, such as TCP. In addition, the protocol mechanisms encapsulated within a microprotocol may be implemented in hardware or software. If appropriate hardware is available, the microprotocol may merely act as a wrapper, calling the relevant hardware function. Like many modular operating systems, Linux supports loadable modules. These software objects encapsulate microprotocols, allowing code to be dynamically loaded into a running operating system and executed without the need to recompile a new kernel. Each such microprotocol can be implemented via a number of adaptable functions, summarized in Table 1.

Whilst a protocol defines the structure and resources available for constructing a communication system, a *protocol stack* defines a unique instantiation assigned to a particular connection. In terms of microprotocols, a protocol stack is an ordered set drawn from some *parent* protocol and combined to form a functional communication system. Each connection is assigned a protocol stack for its sole use, the configuration of which may vary from other stacks derived from the same parent.

Table 1: Adaptable Functionality in DRoPS

Protocol	Implementations
Sequence control	none | complete
flow control	none | window based
Acknowledgement	IRQ | PM-ARQ
Checksums	none | block checking | full CRC

3 Neural network-based Management Scheme for QoS Parameters

In our approach, a Multi-Layer Perceptron (MLP) is used to suggest alternative microprotocols depending on current network conditions (i.e. the monitored QoS). This choice is motivated by the overhead of one order of magnitude smaller compared to other intelligent heuristic approaches and also because the MLP provides generalization capabilities. Although MLPs may not be suited for safety critical applications, the majority of distributed multimedia applications have 'soft' deadlines and the small overheads associated with an MLP can be justified in these cases.

The MLP architecture comprises 10 input nodes, two hidden layers (with fifteen nodes each) and 9 output nodes (denoted by 10-15-15-9 in Figure 1). During execution, end-to-end performance characteristics are collected by each individual microprotocol. Thus, the input to the MLP consists of a vector of 10 values representing measured and error values, appropriately scaled, for the five QoS parameters considered (bit error, segment loss, segment order, delay, jitter). These values are updated dynamically, at run-time. During network operation, the values written in the nodes of the output layer represent the set of functions that should appear in a new protocol configuration. To achieve this, output nodes are grouped logically according to the class of operations they perform; individual nodes represent a single function within that class. Thus, nodes corresponding to the checksum functions represent no error checking, weak error checking (such as that provided by a block checksum), and an alternative strong error check (the Cyclic Redundancy Check). Output nodes may also represent the absence of a particular function, such as the one representing no error checking in our approach. The value in each output node will represent a degree of confidence that the function represented should appear in any new configuration. For example, if error checking is not required, the node representing no error checking will exhibit the highest value in the group of nodes corresponding to the checksum functions.

To reduce processing overhead, only protocol functions that have alternative microprotocols are represented in the output layer. The overheads the MLP have been shown to be small - in the order of 190s - and predictable - it executes with a variance of +/-6s for all cases, lending further support to this choice of adaptation policy mechanism.

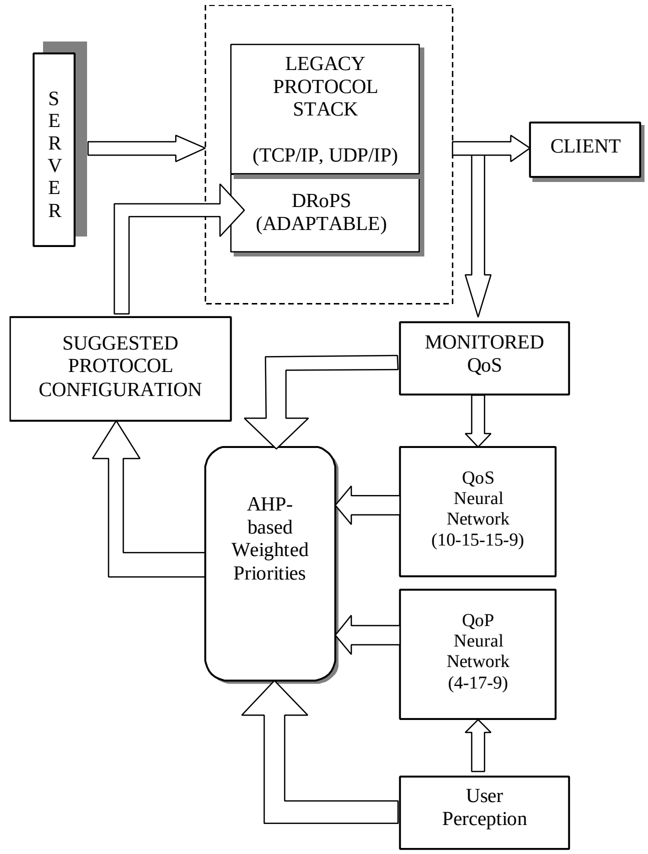

Figure 1: QoP Oriented Management Scheme

4 Neural network-based Management Scheme for QoP Parameters

In this case, an MLP with 4 input nodes, one hidden layer (comprising 17 nodes) and 9 output nodes was used. Here user input can reflect personal choices of the relative importance of the video, audio and textual components in the context of the application, as well as a relative characterization of the dynamism of the multimedia clip. These values can also change depending on the particular scene being visualized. Whilst, these values can be changed dynamically, *a priori* values in this case could reflect the result of consultations with users. This indeed is the case with our QoP experiments, [7], where a broad base of people were polled about their opinions for a wide range of clips with a variety of subject matter (see Table 2). For example, users characterized the News clip as being relatively static, i.e. dynamism is *Low* (L), and considered that the video and audio media components conveyed a *High* (H) informational load, in contrast to the text component which was judged to have a *Medium* (M) informational weight.

Table 2: QoP Video Categories

VIDEO CATEGORY	Dynamic	Audio	Video	Text
1 - Action Movie	H	M	H	L
2 - Animated Movie	M	M	H	L
3 - Band	M	H	M	L
4 - Chorus	L	H	M	L
5 - Commercial	M	H	H	M
6 - Cooking	L	H	H	L
7 - Documentary	M	H	H	L
8 - News	L	H	H	M
9 - Pop Music	M	H	H	H
10 - Rugby	H	M	H	M
11 - Snooker	L	M	M	H
12 - Weather	L	H	H	H

The output of the QoP neural network is interpreted as the output of the QoS neural network described in the previous section.

5 Creating Compromise to Yield Weights Priorities

The neural network-based management schemes for the QoS and QoP parameters suggest alternative protocol configurations, as described in the previous two sections. However, in many cases the fact that the two schemes must satisfy, generally, conflicting objectives and constraints and the considerable uncertainty about the consequences of various alternatives complicate the construction of the appropriately tailored communication protocol.

To alleviate this situation we have used the Analytic Hierarchy Process (AHP), [14], as an additional mechanism that will allow us to handle possible conflicting suggestions of the two neural networks used in our approach. According to the AHP, the main task is to allocate numerical measures to the relative importance of the criteria (QoS and QoP parameters) by taking into account the relative performance of the tailored communication protocols on these criteria.

For this purpose, let us define a_{ij} as the numerical judgement made by the designer in comparing criterion to criterion (i.e. parameter to parameter). These pairwise comparisons are quantified by using a scale. Psychological experiments have shown that individuals cannot simultaneously compare more than 7 objects (±2) [10]. Thus, usually, a scale of 9 grades, which describe the relative importance of the criteria, is suggested (see [14]).

Table 3: Characterizations of the Relative Importance between Activities

Importance	Characterization	Explanation
1	Equally important	Two activities contribute equally to the objective
3	Moderately important	Experience and judgement slightly favour one activity over another
5	Essential or strongly important	Experience and judgement strongly favour one activity over another
7	Demonstrated important	An activity is strongly favoured and its dominance has been demonstrated in practice
9	Absolutely important	The evidence favouring one activity over another is of highest possible order of affirmation
2,4,6,8	Intermediate values	When compromise is needed
Reciprocals of above nonzero	If activity *i* has one of the above nonzero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*	

Table 3, above, presents the basic linguistic terms that can be assigned to a comparison between two activities/criteria, i.e. *"A is equally important as B"*. Intermediate terms can also be assigned when compromise is needed between two adjacent characterizations. Thus, if a_{ij} is a point on this nine-point scale, i.e. $a_{ij} \in \{1,2,3,...,8,9\}$, then $a_{ji} = 1/a_{ij}$ also holds [14],[18]. Finally, a *pairwise comparison matrix* $A = [a_{ij}]$ can be constructed:

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1p} \\ \frac{1}{a_{21}} & 1 & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{p1}} & \frac{1}{a_{p2}} & \cdots & 1 \end{bmatrix}. \qquad (1)$$

Now, let w_1, w_2 ,..., w_p be the weights for the p criteria (QoS and QoP parameters), where w_I denotes the relative importance of the ith criterion. Note that, in order to obtain the preference weights of the criteria, w_i can be evaluated by the equation:

$$w_i = \frac{\left(\prod_{j=1}^{p} a_{ij}\right)^{1/p}}{\sum_{i=1}^{p}\left(\prod_{j=1}^{p} a_{ij}\right)^{1/p}} \quad i = 1,2,..., p \qquad (2)$$

and a higher priority setting corresponds to a greater importance.

At this point we should mention that the consistency of the judgements in a comparison matrix is reassured by means of the Consistency Index (*CI*); the smaller the value of the *CI*, the more consistent are the judgements. Literature suggests that CI<0.1 implies consistency of the judgements [14], [18]. Note that, as stated by Saaty, [14], improving consistency does not mean getting an answer closer to the truth but only that the estimates in the matrix, as a sample collection, are closer to being logically related than to being randomly chosen. In this paper we consider judgements which are consistent, and, thus, MDM problems under certainty. However, this approach can handle inconsistent matrices by exploiting a fuzzy formulation of the problem as suggested in [9].

In our case, following consultations with designers and users, we have obtained a matrix detailing pairwise comparisons between the QoS and QoP parameters:

$$M = \begin{bmatrix} 1 & 1/2 & 1 & 1/4 & 1 & | & 2 & 4 & 4 & 3 \\ 2 & 1 & 1 & 1/4 & 1/3 & | & 5 & 4 & 5 & 4 \\ 1 & 1 & 1 & 1/3 & 1/2 & | & 4 & 6 & 4 & 4 \\ 4 & 4 & 3 & 1 & 5 & | & 6 & 7 & 6 & 5 \\ 1 & 3 & 2 & 1/5 & 1 & | & 4 & 6 & 6 & 6 \\ - & - & - & - & - & + & - & - & - & - \\ 1/2 & 1/5 & 1/4 & 1/6 & 1/4 & | & 1 & 1/2 & 3 & 2 \\ 1/4 & 1/4 & 1/6 & 1/7 & 1/6 & | & 2 & 1 & 3 & 2 \\ 1/4 & 1/5 & 1/4 & 1/6 & 1/6 & | & 1/3 & 1/3 & 1 & 1 \\ 1/3 & 1/4 & 1/4 & 1/5 & 1/6 & | & 1/2 & 1/2 & 1 & 1 \end{bmatrix}. \qquad (3)$$

Here, the respective parameters (criteria) are, in order, bit error (*BER*), segment loss (*SL*), segment order (*SO*), delay (*DEL*) and jitter (*JIT*), Video (*V*), Audio (*A*), Text (*T*) and Dynamism (*D*).

An average user would have difficulty in *a priori* judgment of varying technical parameters such as delay, jitter, error and loss rates on highly subjective attributes, such as perception, understanding and satisfaction. Whilst this is true for QoS attributes at the level of the transport service, users are better able to quantify their requirements in terms of more abstract characteristics like the prioritization of core multimedia components such as video, audio and text.

Thus, the matrix *M* in Relation (3) reflects this situation and could conceptually be split-up into 4 sub-matrices, which are:

- A 5×5 matrix, in the upper left part of matrix *M*, giving the relative importance of the QoS parameters, i.e. *BER*, *SO*, *SL*, *DEL* and *JIT*, with respect to one another.
- A 4×4 matrix, located in the bottom right of the matrix *M* which reflects default perceptual values of the relative importance of QoP parameters, i.e. *V*, *A*, *T* and *D*, with respect to each other for the particular multimedia application viewed. This part of the matrix changes depending on the multimedia category (see Table 2).
- A 5×4 matrix and a 4×5 (one of which is the transpose of the other) which reflect designer choices of the relative importance of the five QoS parameters considered on the QoP parameters. The elements of these matrices remain fixed at run-time, and, in our particular case, reflect the results of our previous work on QoP [7].

Changes in the operating environment, as well as changes in user preferences and perceptions are reflected in our model by periodical monitoring of the current values of the QoS and QoP parameters. Together these 9 parameters make up the elements of a dynamically updated diagonal matrix *U*:

$$U = diag\{BER_p, SO_p, SL_p, DEL_p, JIT_p, V_p, A_p, T_p, D_p\}, \qquad (4)$$

where the subscript p denotes the value of the respective parameter after quantization according to the nine-point scale.

For example, as a result perhaps of a delay-intolerant audio application being subjected to a period of high network delays, the matrix *M* will update its values to:

$$M' = MU \qquad (5)$$

The new matrix M' will ensure that a new set of preference weights, given by (2), is calculated. These reflect a more radical bias in favor of microprotocols which reduce the delay component of QoS.

The existing and suggested configurations are compared and appropriate adaptation commands issued to convert the former into the latter. Protocol functions, drawn from the library of protocol mechanisms given in Table 1, are added, removed and exchanged and the updated protocol configuration used for subsequent communication. The DRoPS runtime framework ensures that changes in protocol configuration are propagated and implemented at all points of communication. The new configuration should then provide a connection with characteristics that match the required performance more closely than the old configuration.

6 Conclusions

In this paper we have applied the multicriteria decision making formalism to obtain a method which, from combined user-, application- and network-level requirements, ultimately results in a protocol configuration specifically tailored for the respective requirements. Our method uses neural network-based QoP and QoS management schemes ensuring dynamic protocol adaptation, as a result of changes in user preferences or the operating environment.

7 References

[1] R.T. Apteker, J.A. Fisher, V.S. Kisimov, and H. Neishlos, "Video Acceptability and Frame Rate", *IEEE Multimedia*, Vol. 2, No. 3, pp. 32-40, 1995.

[2] S. Ardon, R. DeSilva, B. Landfeldt, and A. Seneviratne, "Total Management of Transmissions for the End-User, a Framework for User Control of Application Behaviour", *Proceedings of the 4th International Workshop on H Performance Protocol Architectures (HIPPARCH'98)*, London, June 1998.

[3] J.S. Dyer, P.C. Fishburn, R.E. Steuer, J. Wallenius, and S. Zionts, "Multiple Criteria Decision Making, Multiattribute Utility Theory: The Next Ten Years", *Management Science*, Vol. 38, No.5, pp. 645-654, 1992.

[4] R.S. Fish and R.J. Loader , "A kernel based adaptable protocol architecture", in *Proceedings of the 24th Euromicro Conference*, Västerås, Sweeden, pp. 1029-1036, 1998.

[5] M. Fares and B. Kaminska, "Exploring test space with fuzzy decision making", *IEEE Design and Test of Computers*, pp. 17-27, Fall 1994.

[6] K. Fukuda, N. Wakamiya, M. Murata, and H. Miyahara, "QoS Mapping between User's Preference and Bandwidth Control for Video Transport", in *Proceedings of the 5th International Workshop on QoS (IWQoS)*, New York, USA , May 21-23, pp. 291 – 301, 1997.

[7] G. Ghinea and J.P. Thomas, "QoS Impact on User Perception and Understanding of Multimedia Video Clips", in *Proceedings of ACM Multimedia '98*, Bristol, United Kingdom, pp.49-54, 1998.

[8] G. Ghinea, J.P. Thomas and R.S. Fish, "Quality of Perception to Quality of Service Mapping Using a Dynamically Reconfigurable Communication System", in *Proceedings of IEEE Globecom '99*, Rio de Janeiro, Brazil.

[9] L. Mikhailov, "A fuzzy programming method for deriving priorities in the analytic hierarchy process", *Journal of the Operational Research Society*, pp.341-349, Vol. 51, 2000.

[10] G.A. Miller, "The magical number seven plus or minus two: some limits on our capacity for processing information", *Psychological Review*, Vol. 63, pp. 81-97, 1956.

[11] K. Nahrstedt, K., A. Hossain and S. Kang, "A Probe-based Algorithm for QoS Specification and Adaptation", *Proceedings of the 4th International Workshop on QoS (IWQoS)*, Paris, France, March 6 -8, pp. 115 - 124, 1996.

[12] K. Nahrstedt, "Challenges of Providing End-to-End QoS Guarantees in Networked Multimedia Systems", *ACM Computing Surveys Journal*, December 1995, available online at *http://nahrstedt.cs.uiuc.edu/~klara/ home.html*.

[13] D. Reininger, D. Raychaudhuri and M. Ott, "A Dynamic Quality of Service Framework for Video in Broadband Networks", *IEEE Network*, pp. 22 - 34, November/December 1998.

[14] T. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill International, New York, 1980.

[15] A. Seneviratne, M. Fry, V. Withanan, V. Sapapramdu, A. Richards, and E. Horlait, "Quality of Service Management for Distributed Multimedia Applications", IEEE Communication Society, *Proceedings of the Phoenix Conference on Computer and Communications"*, Phoenix, Arizona, 1994.

[16] R. Steinmetz, "Synchronisation Properties in Multimedia Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 8, No. 3, pp. 401 - 412, April 1990.

[17] R. Steinmetz, "Human Perception of Jitter and Media Synchronisation", *IEEE Journal on Selected Areas in Communications*, Vol. 14, No. 1, pp. 61 - 72, 1996.

[18] E. Triantaphyllou E. and S. Mann, "Using the analytic hierarchy process for decision making in engineering applications: some challenges", *International Journal of Industrial Engineering: Applications and Practice*, Vol. 2, No. 1, pp. 35-44, 1995.

[19] A. Watson and M.A. Sasse, "Measuring Perceived Quality of Speech and Video in Multimedia Conferencing Applications", *Proceedings of ACM Multimedia '98*, Bristol, U.K., pp. 55 - 60, 1998.

[20] D. Wijesekera and J. Srivastava, "Quality of Service (QoS) Metrics for Continuous Media", *Multimedia Tools and Applications*, Vol. 3, No. 1, pp. 127 - 166, September 1996.

[21] D. Wijesekera, J. Srivastava, A. Nerode and M. Foresti, "Experimental Evaluation of Loss Perception in Continuous Media", available online at *http://www-users.cs.umn.edu/~wijesek/ publications/pub.html*.