QUALITY-DISTORTION THEORY: THE OPTIMAL LEVEL OF DISTORTION WHEN FIDELITY IS EXPENSIVE

Virgilio Rodriguez

ECE Department, Polytechnic University
5 Metrotech Center, Brooklyn, NY 11201
E-mail:vr@ieee.org

ABSTRACT

Below, 2 related problems are analyzed, by postulating that a quality/distortion (Q-D) curve (or "utility function") describes how a human perceives an "imperfect" media signal. First, a consumer with a limited budget can acquire more media files, by accepting more distortion per file. The amount of distortion that maximizes *the sum* of the utility of each purchased file is found and clearly identified in the Q-D curve. Likewise, an energy-limited transmitter with many media files to transfer can, statistically, reduce distortion per file, at the expense of fewer transferred files. A solution that maximizes his *total expected* utility is given as a specific point in the graph of the *expected* utility as a function of the received SIR. This formulation is very robust, because the proposed family of Q-D curves contains as a special case the step function implicitly assumed by the literature, as well as many other plausible shapes.

KEY WORDS: rate/distortion theory; mathematical modeling; resource management; multimedia communication; power control

1. INTRODUCTION

Distortion measures the difference between a signal and its copy. It is an important QoS measure in the processing and transmission of error-tolerant information, such as media signals intended for human consumption. Typically, when dealing with distortion, the resource-management literature assumes that up to a level, distortion is of no consequence, but beyond that level, it makes the signal totally useless. Such "hard threshold" seems at odds with the way humans process media signals. These signals can be useful at various degrees of noticeable distortion. And when a reduction of distortion is costly, the consumer can prefer more distortion, in exchange for energy, money, or other savings. Furthermore, scientific work has shown that judiciously relaxing the distortion constraint by a small amount can lead, under certain conditions, to a disproportionately larger increase in the capacity of a CDMA network[1].

Hence, a tractable model is needed for the way humans perceive the quality of "imperfect" signals. Below, a model that establishes a quality-distortion relation is introduced. The model is sufficiently flexible to capture a wide variety of plausible quality-distortion relationships, and includes as special cases some of the simpler cases, such as the step function often assumed by the literature. It is postulated that the perceptual quality of an imperfect copy of a signal is determined by a sensible decreasing function of its distortion. No specific algebraic functional form ("equation") is imposed. Rather, a general family of Q-D functions is assumed. Any such function has the general shape shown in fig. 2. This shape can accommodate a wide variety of quality-distortion relations ("step", "ramp", convex, concave, etc).

Generally, the literature assumes that distortion has no noticeable effect up to a certain level, and completely spoils the signal after that level. Reference [2] takes a somewhat more general approach by postulating that the end-user wishes to maximize the "utility" of an imperfect media signal. But this reference focuses on video over a wired network, and only considers the special case of a logarithmic utility function. Also of interest is [3], which consider the wireless transmission of images that have been scalably encoded, as with the JPEG-2000 standard. This references postulates that the quality ("utility") of the image resulting when a truncated scalable file is decoded is given by an increasing S-curve defined on the number of bits in the truncated file (coding rate). Because the family of Q-D curves ("utility functions") assumed in the present paper includes as special cases both the logarithmic and the step function, the present approach is a strict generalization of the literature.

Under this approach, the "right amount" of distortion is a variable to be chosen optimally, whether directly, or, by choosing other resources, indirectly. Below, a situation in which distortion is directly chosen is considered first. A consumer is offered media files at various degrees of distortion. Both his "utility" and the cost of acquiring a file are

Supported in part by NYSTAR through the Wireless Internet Center for Advanced Technology (http://wicat.poly.edu) at Polytechnic Univ.

decreasing in the amount of distortion in the file. With a limited budget, which could be in money, energy, time or any other valuable resource, the consumer faces a classical quantity vs. quality trade-off. He can obtain relatively few high-quality media files, or relatively many low-quality ones. What is the optimal choice? It turns out that with linear pricing the optimal amount of distortion can be quite clearly described. It is obtained by drawing a tangent line from the point $(0, \bar{D})$ to the graph of the utility function ($\bar{D}$ is the largest available distortion level). With non-linear pricing, a similar but somewhat more involved procedure can be applied.

A more specific communication scenario is also considered. An energy-limited transmitter with many media files (images) to transfer over a wireless link wants to choose optimally its transmission power. At low transmission power, many bit errors occur, which produce a highly distorted image at the receiver. High transmission power produces less distortion, at the expense of higher energy consumption per file. Again, a quality vs. quantity trade-off arises. The transmitter opts to maximize the total *weighted* number of files transferred before energy runs out. The weight of each file is its expected "utility" (perceptual quality), which is a function of its distortion. This distortion is a random variable determined by the number of bit errors during the transmission of the file, which is itself determined by the signal-to-interference ratio (SIR), γ, at the receiver. With $\bar{U}(\gamma)$ denoting the expected utility of a media file, the analysis leads to choosing an SIR γ^* to maximize an index in utility/Joule, which is proportional to $\bar{U}(\gamma)/\gamma$. For bit-error functions of practical interest, $\bar{U}(\gamma)$ has the familiar S shape, and γ^* can be obtained by drawing a tangent line from the origin to the graph of $\bar{U}(\gamma)$ (see x^* in fig. 3)[4, 5].

Below, the general properties of the proposed family of Q-D curves are formally given and discussed. Then, the situation in which the degree of distortion of media files can be directly chosen optimally given a cost function is analyzed. Subsequently, the more specific telecommunication problem is solved. Finally, some general summarizing comments are given. (Below, the phrase "perceptual quality" and "utility" are used exchangeably. Strictly speaking, a difference could be established between the two)

2. QUALITY/DISTORTION THEORY

Distortion is typically defined as a relatively simple mean square measure of the difference between a signal and its copy. As an indicator of media quality as perceived by a human observer, this index is, at best, a very crude measure. The *perceptual* quality of an "imperfect" copy of a signal is determined by the human sensory system (visual, auditory, etc). It seems reasonable to assume that the perceptual quality is somehow determined by distortion; i.e., that a function $Q(D)$ that translates distortion into perceptual quality can be found. The quality-distortion function cannot be derived, and should not be imposed. It should be obtained by psychophysical experimentation. However, one can make some reasonable assumptions about the properties that any such function should possess. Then one can analyze a problem of interest and (optimistically) describe its solution by employing the general properties of the curve.

2.1. Intuitive specification

Figure 1 illustrates some plausible, simple $Q(D)$ curves. First is, of course, the supposition that perceptual quality falls linearly as distortion increases from zero to its highest value ("quality equals fidelity"). This assumption would greatly simplify the analysis. But it essentially means that the human visual system (HVS) (or auditory, etc) is perfectly "tuned" to a very simple mean squared measure, ..., in all cases, ..., for all people. Such a strong assumption would be adventurous, and likely to be refuted by experimentation. Another highly simplifying assumption often employed in the literature is that distortion is unnoticeable up to a level (c in fig. 1) but it totally spoils the signal beyond that point ($Q(D)$ is a "step function"). But our own experience tells us that media signals can be useful at various degrees of noticeable distortion. Furthermore, when a reduction of distortion is costly, a human may choose to tolerate more distortion, in exchange for energy, money or other savings. But the step function assumption precludes the study of such trade-offs. A third possibility illustrated in fig. 1 is the "ramp" $Q(D)$, implying that distortion has no noticeable effect up to a level (a), and completely spoils the signal beyond another level (b), while varying linearly between these two points. Presumably, a and b would be determined by the specific user/application combination. The ramp includes as special case the threshold ($a = b = c$) and the linear relation ($a = 0$, $b = D_{\text{MAX}}$); but still its "piecewise linearity" is a big imposition which may not be supported by experimentation.

Further reflection indicates that it is reasonable to assume that the graph of the $Q(D)$ function is a "reversed" S-curve, as shown by fig. 2. This graph strictly generalizes the step function often assumed in the literature. And the family of S-curves includes as special cases curves that are "mostly" convex, others that are "mostly" concave, and some whose "ramps" follow closely a straight line over a given interval. Thus, if the analyst assumes that *all that is known* about the $Q(D)$ curve is that it is a reverse S-curve, and conducts the analysis on the basis of properties derived from this shape, the solution procedure and conclusions will be valid for a wide variety of plausible $Q(D)$ relations.

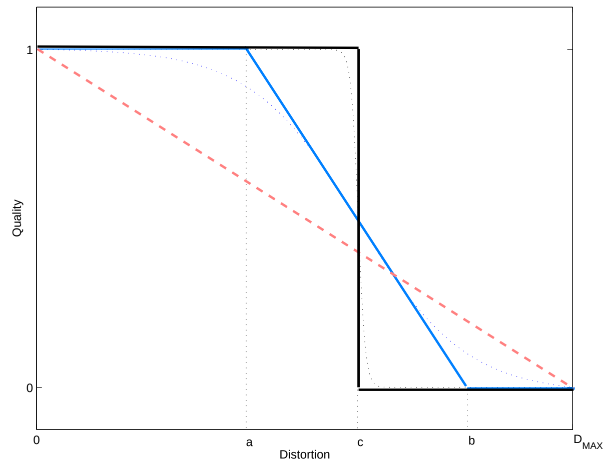

Fig. 1. Quality vs. distortion: Some plausible simple relations are: (i) fidelity equals quality (red dashed line); (ii) hard threshold (step); (iii) ramp (blue broken line). The ramp includes as special case the threshold ($a = b = c$) and the linear relation ($a = 0$, $b = D_{\text{MAX}}$). But the reverse S-curve includes all these cases and more (see next figure).

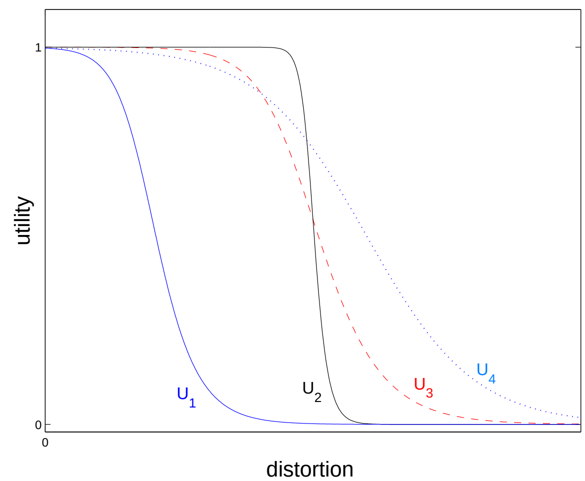

Fig. 2. In the eyes of the beholder: Media signals can be useful to end users at various degrees of noticeable distortion. This is captured by a "utility function" indicating the "usefulness" of the distorted signal.

2.2. Formal definition

The Q-D curve ("utility function") has the following properties:

1) Its domain is the interval $[0, \bar{D}]$, where $\bar{D}$ is the largest available level of distortion.

2) Its range is the interval $[0, 1]$. This is just a normalization. A 1 denotes the best possible quality of the decoded file (say the quality of the original) and a zero is the 'quality' of a maximally distorted file .

3) It is strictly decreasing (distortion worsens quality)

4) Its graph is "reversed" S-shaped, as in fig. 2. In practical terms, reversed-sigmoidness further implies that: (a) If the distortion is sufficiently small, the quality of the decoded file will be sufficiently close to "perfect". (b) After distortion has been sufficiently reduced, the marginal contribution to media quality of further reductions of distortion becomes "very small" and is decreasing. (c) If the distortion is sufficiently large, the quality of the decoded image will be sufficiently close to zero. (d) The function becomes convex as distortion increases ("eventual convexity"). One plausible interpretation is that even a highly distorted image may provide enough information to identify its "meaning" (what is it? a bird?, a person's face?, etc.). This essential semantic information is provided at high levels of distortion. Thus, the utility of the distorted image *increases at a fast rate* as distortion is *reduced from its highest level* (right to left in the graph).

2.3. An alternate view: fidelity vs. distortion

Rather than basing the argument on the distortion, y, of the recovered signal, one can focus on the variable $x = \bar{D} - y$, interpreted as the "fidelity", or the amount of distortion which has been "avoided" or "removed". When $x = 0$ the resulting signal is "fully" distorted ($y = \bar{D}$). We can think of this as a signal obtained by guessing all the bits in the concerned file, which yields the "cheapest" possible image. To get an image with any less distortion necessitates some kind of expenditure. The larger x (the difference between $\bar{D}$ and y), the higher the quality of the image, and the greater its cost. Thus, this analysis can be based on the derived function $s(x) := u(\bar{D} - y)$. The graph of $u(-y)$ is the "mirror image" of that of u ("time reversal"). And the graph $u(\bar{D} - y)$ is the same as that of $u(-y)$ but shifted to the right $\bar{D}$ units. Thus, the graph $s(x)$ yields a "standard" S-curve, as displayed in fig. 3. This observation will prove useful in the technical development.

3. ACQUIRING VARIABLY DISTORTED INFORMATION

Pedagogically, it may be useful to set up the problem of interest in a general scenario, before introducing communication issues.

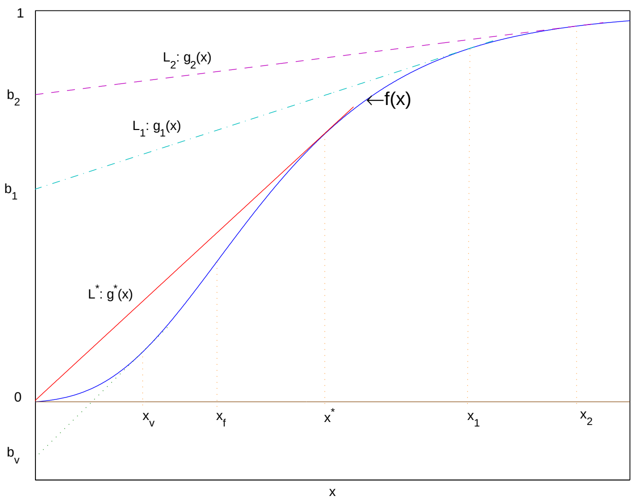

Fig. 3. An “S-curve” and some of its tangents

3.1. Problem statement

A consumer can acquire files corresponding to perceivable media (say images), each available at varied degrees of distortion, $y \in [0, \bar{D}]$. The cost of any one image (in terms of money, energy, or any other scarce resource that the consumer has and values) is $c(y)$, which is always positive and decreasing in the level of distortion, y. For convenience, let $c(\bar{D}) = 0$ and $c(0) = c_0$. Images are equally valuable to the consumer, in the sense that he is indifferent between any one of two images, if they both have the same level of distortion. The usefulness, quality, or “utility” to the consumer of a distorted image is determined as a function of its distortion, y, by a function $u(y)$, whose properties are discussed in section 2.2.

The consumer wants to spend his budget B optimally. That is, he wants to determine, given u, c and B, what is the “right” amount of distortion he should choose. If he chooses to acquire images with very small distortion ($y \approx 0$), the cost of each image, $c(y)$, will be “high”, and the number of images he will get to view, $B \div c(y)$, will be small. On the other hand, choosing a large y will result in a large number of highly distorted images.

Notice that, as discussed in section 2.3, the problem can be stated in terms of $x = \bar{D} - y$, which is interpreted as the amount of distortion which has been “avoided” or “removed from” the image, or simply its “fidelity”. In this case, the pertinent cost function is denoted as $c_x(x)$.

3.2. Objective Function and Constraints

Some reflection indicates that the consumer should maximize his total utility, which is obtained as the product of the quality (or utility) of each image by the total number of images he gets acquire. Hence, the consumer should solve

$$\max_{0 \le y \le \bar{D}} \quad \frac{u(y)}{c(y)} \tag{1}$$

$$\text{or} \quad \max_{0 \le x \le \bar{D}} \quad \frac{s(x)}{c_x(x)} \tag{2}$$

(multiplying by the constant B would make no difference to the solution).

Now, the index being maximized, $u(y)/c(y)$ or $s(x)/c_x(x)$, has the unit quality/dollar, quality per Joule, or quality per second, depending upon the customer's scarce resource.

3.3. First-order optimizing conditions

The first-order necessary conditions (FONOC) for an interior solution to this problem is

$$c(y)u'(y) = c'(y)u(y) \tag{3}$$

$$\text{or } c_x(x)s'(x) = c'_x(x)s(x) \tag{4}$$

Inspection of this equation immediately indicates that if $c(y) \propto u(y)$ then any value of y (or x) would satisfy it.

3.4. Solutions

3.4.1. Linear cost function

If c is such that $c(y) = (\bar{D} - y)\bar{c}$, ($c_x(x) = \bar{c}x$) then the objective function (eq. (3)) can be written, as

$$\max_{0 \le x \le \bar{D}} \frac{u(D-x)}{x} = \frac{s(x)}{x} \tag{5}$$

As discussed in section 2.3, the graph of $s(x)$ has the form shown in fig. 3; that is, $s(x)$ is a standard “S-curve”. The solution to maximizing $s(x)/x$, with s an S-curve, is well understood[5]. It is the unique positive number obtained as the abscissa of the point at which a tangent line emanating from the origin meets the graph of s. (see x^* in fig. 3). The optimal distortion level is $y^* = \bar{D} - x^*$. Equivalently, the desired solution can be obtained by drawing a tangent from the point $(\bar{D}, 0)$ to the graph of the original $u(y)$.

3.4.2. General convex cost functions

The preceding development can be extended, with due attention to certain technical details, to an arbitrary cost function, via a non-linear change of variable. Further details can be obtained from the author.

4. DISTORTION AND POWER MANAGEMENT

Below, the analysis focuses on the more specific scenario of transmission of error-tolerant files (“media”) over a wireless link. For simplicity, each information bit in a file is viewed as corresponding to a pixel of an *uncoded* black and white image.

4.1. Problem statement

It is taken as given: (1) a certain amount of energy, E, available for transmission; (2) a fixed transmission rate of R bits per second; (3) a long sequence of files, each corresponding to an equally important image, and each divided into N blocks of bits (packets) with a total of M bits, of which L are information bits; (4) a certain level of interference (noise), I. The transmission proceeds one packet at a time, *without* retransmissions. An error-control system is *assumed* to operate as follows. Up to m bit errors per packet can be corrected; and if $m+1 \leq k \leq L$ bit errors occur in a given packet, each will ultimately contribute one error in the decoded file. These errors creates distortion. Thus, there is also a function u as defined in section 2 giving the utility (quality) of a received file as a function of its distortion.

The signal-to-interference ratio (SIR) at the receiver determines the bit error probability. Thus, a larger transmission power leads to fewer errors, statistically lower values of distortion, and greater *expected* utility. But, with limited energy, more transmission power means fewer total images transferred. The transmitter wishes to utilize its energy efficiently.

4.2. Distortion analysis

The error-control system is viewed as a "black box" whose net effect is that a packet with $m+1 \leq k \leq L$ bit errors contribute k errors to the decoded file. Distortion is, generally, defined as of sum of squares of differences between the reconstructed signal and the original. This sum equals the total number of bit errors in the reconstructed image, in this scenario.

For example, suppose that the number of packets per file is 2, and that the code being used can correct up to 3 bit errors per packet. Suppose that 2 and 5 bit errors occur during the transmission of the first and second packet, respectively. Then, the first packet is corrected, so that all its information bits coincide with the original. But the 5 errors in the second and final packet are not corrected, and contribute 5 errors in the decoded file. Thus, the total distortion of this image will be 5. The utility function of the user will determine how good or bad a distortion of 5 is.

It is worth noting that it is not obvious, at least in this problem, what is the worst case scenario for distortion. In principle, it would seem that having each and every bit in error should be the worst that can happen. However, given the idiosyncrasies of the human visual system, if a black and white image were to have each and every bit reversed, the result would be a perfectly intelligible image, in which black and white simply switch roles! However, this fact is not considered in the analysis below.

4.3. Expected utility of distorted image

In this scenario, distortion is a discrete random variable. The transmission power determines the bit-error rate (BER), and frame-error rate (FER), which indirectly determines the probability distribution of distortion. When the number of packets per file , N, is large, expressing this probability distribution in terms of the BER is quite cumbersome and tedious. This task is, however, relatively straightforward when each image fits into a single packet. Let this be the case. Under the assumptions that have been made about the error-control system, distortion is zero, if m or less bit errors have occurred during the transmission of the packet. When the number of bit errors exceed the number that can be corrected by the code, what happens depends on more specific details of the error control system. Let us assume, pessimistically, that if $m+1$ to L errors occur, each will cause an error among information bits in the decoded file.

Assuming independent bit errors, the probability of k bit errors in an L bit packet is given by $\binom{M}{k}\epsilon^k(1-\epsilon)^{M-k}$, with ϵ the bit-error rate (BER) which is determined byγ, the signal-to-interference ratio (SIR) at the receiver.

For the single-packet file, given eq. (**??**), the ***expected utility*** of a file $U_E(\gamma)$ is

$$u(0)\underbrace{\left(\sum_{k=0}^{m}\binom{M}{k}\epsilon^k(1-\epsilon)^{M-k}\right)}_{\text{Prob of 0 to m bit errors}} \quad + \\ \sum_{k=m+1}^{L}\binom{M}{k}\epsilon^k(1-\epsilon)^{M-k}u(k) \qquad (6)$$

4.4. Solution

The expected utility function $U_E(\gamma)$ is a representative measure of the expected quality of each image, given a transmission power level, P, which determines the received SIR, γ. Notice, however, that the BER is 1/2 when $\gamma = 0$, which means that $U_E(0) > 0$. To avoid technical problems involving "transmissions" with 0 power, $\bar{U}(\gamma)$:=$\bar{U}(\gamma)$-$\bar{U}(0)$, the "earned" expected utility of an image, is chosen as the representative quality figure of merit (see [4] for a relevant discussion involving error-intolerant data transmissions).

Since each bit lasts $1/R$ secs., (R is the transmission bit rate), the total energy consumed by the transmission of the single-packet image is PM/R. Thus, $ER \div MP$ images can be transferred with E Joules. The transmitter wishes to maximize its total (earned) expected utility, and must solve:

$$\max_{0\leq P\leq \bar{P}} \frac{R}{M}\frac{\bar{U}(\gamma)}{P} \equiv \max_{0\leq\gamma\leq\bar{\gamma}} \frac{R_c}{M}\frac{h}{I}\frac{\bar{U}(\gamma)}{\gamma}$$

with h the path loss, I the interference, R_c the "chip rate" (a CDMA constant closely related to the bandwidth), $\bar{P}$ the highest available transmission power, and $\bar{\gamma} = (R_c/R)h\bar{P}/I$, the highest achievable SIR.

It can be argued that for BER functions of practical interest, the graph of $\bar{U}(\gamma)$ has the S shape displayed in fig. 3. Then, by the argument given in section 3.4.1, the value γ^* which maximizes $\bar{U}(\gamma)/\gamma$ can be obtained by drawing a tangent from the origin to the graph of $\bar{U}(\gamma)$. This value determines the transmission power, and solves the single user problem. In any case, it is discussed in [4] that if $\bar{U}(\gamma)$ was convex, the optimal would occur at the highest available power level, and that if $\bar{U}(\gamma)$ was concave it would be optimal to "operate" at zero power.

The preceding development also applies when each media file is divided into many packets. Extending the preceding analysis to consider a multi-packet image file is conceptually simple, but very tedious. The procedure to find the probability distribution of distortion is more cumbersome. But once done, it is straightforward to find the "earned" expected utility of a file as a function of the received SIR, $\bar{U}(\gamma)$. The shape of the graph of this function should not be affected by the number of packets per file.

5. DISCUSSION

Media signals can be useful at various degrees of distortion. A proposed model captures this fact mathematically, and enables its exploitation, when avoiding/reducing distortion requires the expenditure of limited resources. Two interesting problems involving a quality versus quantity tradeoff are formulated and solved. In one case, media files are offered at various degrees of distortion, at a price that is *decreasing* in distortion. A consumer willing to accept a higher degree of distortion, can acquire more files. A more specific version of this problem involves an energy-limited transmitter wishing to transfer many images over a wireless link. Spending more energy per packet reduces bit errors, and hence distortion, but also leads to fewer images transferred.

At the core is a function relating the perceptual quality ("utility") of an "imperfect" media signal to its distortion; i.e., a quality-distortion (Q-D) curve. In the development, no specific "equation" (logarithmic, logistic, etc) is imposed as a Q-D function. Rather, it is assumed that *all that is known* about this curve is that it belongs to certain family characterized by a "reversed" S-shaped graph. The analysis follows from the general properties of this family; so that it applies to *any* Q-D curve, as long as its graph has the assumed shape. This shape contains as special case the "sharp threshold" (step) often assumed in the literature, as well as many plausible Q-D relations (convex, concave, "ramps", etc). This level of generality is important, because the "true" Q-D curve can only be obtained by psychophysical experimentation with human subjects. The actual curve will, generally, depend on the specific targeted human user, and quite possibly on the specific application. Because of its generality, this analysis and its conclusions are robust, and should hold for many user/application combinations.

6. REFERENCES

[1] A. Kwasinski and N. Farvardin, "Resource allocation for cdma networks based on real-time source rate adaptation," *IEEE ICC*, vol. 5, pp. 3307 – 3311, May 2003.

[2] C. Luna and A. Katsaggelos, "Maximizing user utility in video streaming applications," *IEEE ICASSP*, vol. 3, pp. 1465 –1468, 2001.

[3] V. Rodriguez, "Resource management for scalably encoded information: The case of image transmission over wireless networks," *IEEE ICME*, vol. 1, pp. 813 –816, 2003.

[4] V. Rodriguez, "Robust modeling and analysis for wireless data resource management," *IEEE WCNC*, vol. 2, pp. 717–722, 2003.

[5] V. Rodriguez, "An analytical foundation for resource management in wireless communications," *IEEE Globecom*, vol. 2, pp. 898–902, December 2003.