Rate, Distortion and Quality: A Tractable Abstraction of the Human Sensory System

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

***Abstract*— Distortion is defined as a relatively simple mean square measure of the difference between a signal and its copy. This index is reasonable and can be useful in many situations, but is, at best, a very crude indicator of media quality as perceived by a human. An abstraction of the human visual (more generally, sensory) system, which is flexible, tractable, and consistent with rate distortion theory, is introduced. Applications are mentioned.**

I. Introduction

Typically, when dealing with distortion, the literature assumes that up to a level, distortion is of no consequence, but beyond that level, it makes the signal totally useless. That is, the literature implicitly assumes that a "step function" relates the quality or "utility" of a reconstructed signal as a function of its distortion. But media signals can be useful to humans at various degrees of noticeable distortion. Thus, the step function does not capture our own experience. Furthermore, the step function assumption precludes the study of some legitimate engineering trade-offs. When a reduction of distortion is costly, a human may choose to tolerate more distortion, in exchange for energy, money or other savings. Furthermore, [1] reports 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. Thus, a tractable model is needed for the way humans perceive the quality of "imperfect" signals.

Furthermore, certain applications requires a clear understanding of the way the coding rate influences perceptual quality. For example, one may be interested in finding an optimal coding rate. This operation is particularly clear when the media files are scalably encoded (JPEG 2000, MPEG 4, etc.). In this case, numerous coding rates can be obtained simply by truncating the encoded file. Presumably, the greater the coding rate the greater the resource expenditure. In order to choose where to truncate one needs to model how the resulting media quality varies with the chosen rate. For instance, [2] discusses the optimal choice of coding rate and transmission power for the wireless transfer of scalably-encoded image files. This analysis is extended in [3] to video streaming. These references postulate that the quality ("utility") of the image/video segment resulting when a truncated scalable file is decoded is given by an S-curve defined on the number of bits in the truncated file. Reference [2] provides some justification for the S-curve, but establishes no connection between the assumed curve and rate-distortion theory.

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. Subsequently, this model yields a rate-quality relationship, that is consistent with rate-distortion theory.

II. Quality/Distortion Theory

Distortion is typically defined as a relatively simple mean square measure of the difference between a signal and its copy. This index can be useful in certain situations. However, as an indicator of media quality as perceived by a human observer, it 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.

Some reflection indicates that it is reasonable to assume that the graph of the $Q(D)$ function is a "reversed" S-curve, as shown by figure 1. This graph strictly generalizes the step function. And the family of S-curves includes as special cases the "step" often assumed in the literature, as well as curves that are "mostly" convex, others that are "mostly" concave, and some whose "ramps" follow closely a straight line.

Of special notice is the fact that a reverse S-curve becomes convex as distortion increases ("eventual convexity"). One plausible interpretation is that even a highly distorted media 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 media *increases at a fast rate* as distortion is *reduced from its highest level*. This model is applied to solve two specific engineering problems in a separate work [5]. Reference [4] discusses the technical characterization of a generic S-curve.

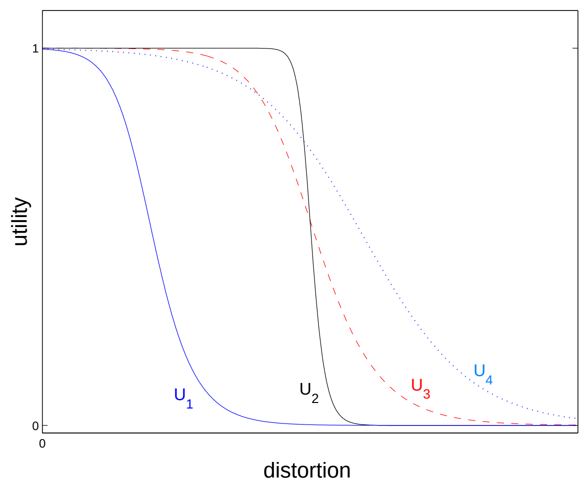

Fig. 1. Perceptual quality ("utility") as a function of distortion.

III. Rate/Quality Theory

When the quality-distortion curve is a "reversed" (decreasing) S-curve, one can characterize the coding-rate-quality relationship by applying basic rate-distortion theory, as illustrated in figure 2. It is generally accepted that the function $D(R)$, giving distortion as a function of the coding rate, is decreasing and convex. With $Q(D)$ denoting the reversed S-curve giving *perceptual* quality as function of distortion, it is clear that the composite function $Q(D(R)) := U(R)$ yields perceptual quality directly as a function of the coding rate. It is then of interest to characterize the composite function $Q(D(R))$ when *all that is known* about $D(R)$ is that it is decreasing and convex, and *all that is known* about $Q(D)$ is that it is a "reversed" S-curve.

The caption of figure 2 contains an approximate analysis that suggests that the graph of $U(R) = Q(D(R))$ is a (non-reversed) S-curve. Figure 3 confirms this conjecture for specific $Q(D)$ curves, and the $D(R)$ function of the memoryless Gaussian source. More analytical work is needed to establish conclusively the properties of the quality-rate curve.

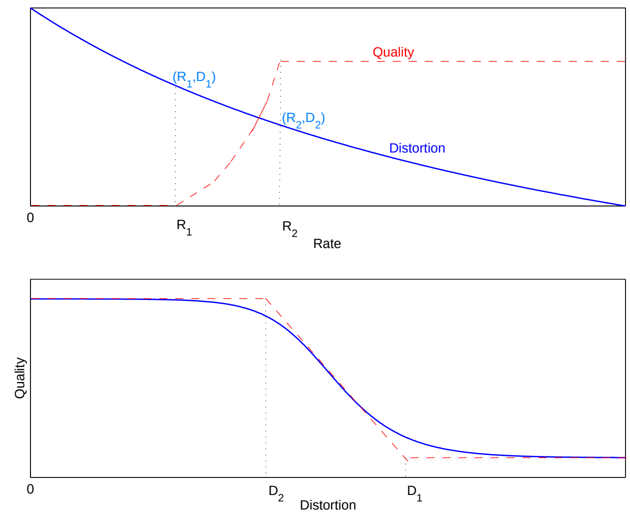

Fig. 2. The convex curve at the top, $D(R)$, gives distortion as a function of the coding rate. The reversed S-curve at the bottom, $Q(D)$, relates *perceptual* quality to distortion. The composite function $Q(D(R)) := U(R)$ yields perceptual quality directly as a function of the coding rate. If $Q(D)$ is approximated by the broken red line at the bottom, the resulting $U(R)$ is the broken red line at the top. Without the approximation, $Q(D(R))$ can be expected to yield an (increasing) S-curve.

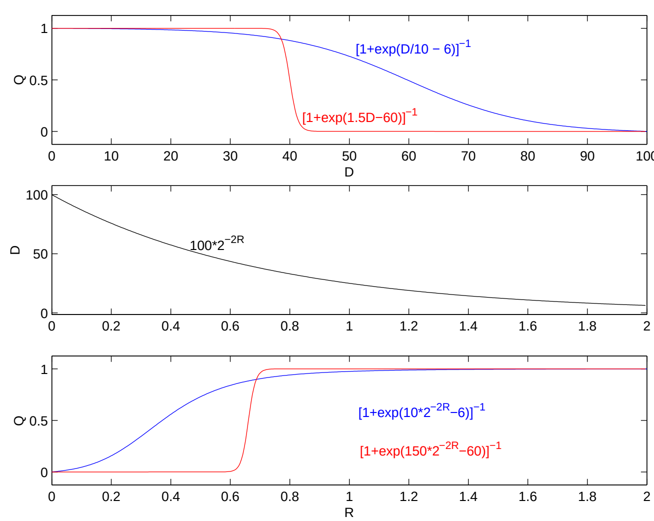

Fig. 3. Two plausible quality-distortion curves, $Q(D)$ (top), with the corresponding $Q(D(R))$ (bottom), for the Gaussian $D(R) \propto 2^{-2R}$

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

[3] V. Rodriguez, "Optimal coding rate and power allocation for the streaming of scalably encoded video over a wireless link," *IEEE ICASSP*, May 2004. (to appear).

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

[5] V. Rodriguez, "Resource management under elastic distortion constraints: Transmission of media files over a wireless link," 2004. Available at http://pages.poly.edu/~vrodri01/research.