NTIA Technical Memorandum TM-03-398

Compensating for System Gain– Motivations, Derivations and Relations for Three Common Solutions

Stephen D. Voran

U.S. DEPARTMENT OF COMMERCE
Donald L. Evans, Secretary

Nancy J. Victory, Assistant Secretary
for Communications and Information

October 2002

CONTENTS

COMPENSATING FOR SYSTEM GAIN: MOTIVATIONS, DERIVATIONS, AND RELATIONS FOR THREE COMMON SOLUTIONS

Stephen D. Voran[1]

It is often desirable to compensate for system gain, especially before objectively estimating perceived audio or video quality from system inputs and outputs. A common approach is to scale the system output to compensate for system gain. One can take three views of the system, and this leads to three different gain compensation solutions: one that minimizes distortion, one that matches input-output power, and one that maximizes signal-to-distortion ratio. We derive these three solutions, describe the algebraic and geometric relationships between them, and provide a generalized result that subsumes all three. We provide audio and video examples and show that these three solutions can differ significantly. We also report some of the gain compensation choices found in the objective audio and video quality estimation literature.

Key words: audio quality estimation, gain compensation, gain estimation, speech quality estimation, system gain, video quality estimation

1. INTRODUCTION

There are numerous engineering situations where it is desirable to compensate for system gain. Two important examples are the objective estimation of perceived audio or video quality [1]-[9]. In general, an audio or video system will distort the audio or video signal and will also apply some non-unity gain factor to the signal. Constant non-unity system gain results in a shift of volume, contrast, or color saturation. Such shifts are often not considered to be part of the distortion introduced by the audio or video system since users often routinely adjust volume, contrast, brightness, and color balance to suit individual preferences. Objective estimators of perceived audio and video quality are intended to give results that agree with human perception. If system gain is not considered to be a distortion component by human subjects, then it should be likewise ignored by objective estimators. This means that the system gain must be compensated for so that the objective estimator can properly measure the actual distortion components. Figure 1 describes the typical approach where the system output $\boldsymbol{y}$ is scaled by the reciprocal of the estimated system gain $\tilde{g}$ to produce a gain-compensated output signal $\tilde{\boldsymbol{y}}$.

[1] The author is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80305.

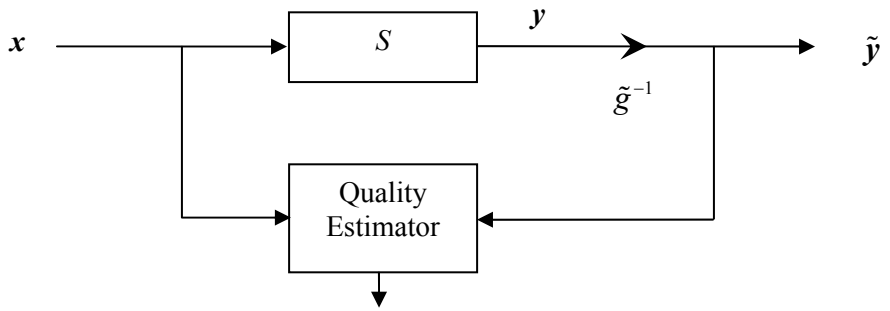

Figure 1. Typical block diagram for gain-compensated quality estimation of the system *S*.

When estimating audio and video quality, there may be other fixed "distortion" components that one might wish to compensate for in order to best emulate human perception. These may include a shift in mean signal level, a constant delay, or small constant spatial shifts of video frames. Compensation for these "distortion" components is outside the scope of this memorandum.

In this memorandum we point out that there are three distinct, mathematically-motivated solutions for compensating for system gain. We derive each solution, describe the algebraic and geometric relationships between them, and provide a generalized result that subsumes all three. Finally we offer example results taken from a digital speech codec and an analog video recorder.

Each of these three solutions for system gain compensation has been used before. Different authors have chosen different solutions and have offered no comment regarding the other choices. Each of the solutions is quite simple and when presented in isolation from the other two, each solution may initially appear to be the only logical choice, thus creating the appearance that no discussion of the solution is necessary. This memorandum shows that discussion *is* warranted because each solution has a unique mathematical motivation, each solves a unique problem, and each yields a unique result. The solution one uses should reflect a conscious choice based on the system under consideration. This memorandum also shows how the three solutions are related to each other.

2. THREE GAIN COMPENSATION SOLUTIONS

In general, we wish to treat the case where the system *S* is not well-modeled by gain plus additive noise. Rather we assume that *S* induces some arbitrary distortion and gain *g* on signal vector $\boldsymbol{x}$ to create the output signal vector $\boldsymbol{y}$. Given a single input vector $\boldsymbol{x}$ and output vector $\boldsymbol{y}$ we must find a reasonable value of $\tilde{g}$ so that scaling $\boldsymbol{y}$ by $\tilde{g}^{-1}$ will give the compensated output signal vector $\tilde{\boldsymbol{y}}$. Without any loss of generality we assume that all signals have zero mean and non-zero magnitude and we assume that the $\boldsymbol{x}$ and $\boldsymbol{y}$ are not orthogonal.

It is certainly reasonable to seek a value of $\tilde{g}$ such that remaining system distortion is minimized:

$$\min_{\tilde{g}} \left| \boldsymbol{x} - \tilde{\boldsymbol{y}} \right|^2 \;\Rightarrow\; \min_{\tilde{g}} \left| \boldsymbol{x} - \tilde{g}^{-1} \boldsymbol{y} \right|^2 \;\Rightarrow\; \tilde{g}_{MD} = \frac{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{y}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}}, \tag{1}$$

where we have used conventional least squares to solve for the minimum distortion gain estimate $\tilde{g}_{MD}$.

For many systems absolute distortion values are less relevant than signal-to-distortion ratios. The nature of human auditory and visual perception makes this especially true for audio and video signals. Thus it would also be reasonable to seek a value of $\tilde{g}$ such that remaining system signal-to-distortion ratio (SDR) is maximized:

$$\max_{\tilde{g}} 10 \log_{10} \left(\frac{\left| \tilde{\boldsymbol{y}} \right|^2}{\left| \boldsymbol{x} - \tilde{\boldsymbol{y}} \right|^2} \right) \;\Rightarrow\; \max_{\tilde{g}} \frac{\left| \boldsymbol{y} \right|^2}{\left| \tilde{g} \boldsymbol{x} - \boldsymbol{y} \right|^2} \Rightarrow\; \min_{\tilde{g}} \left| \tilde{g} \boldsymbol{x} - \boldsymbol{y} \right|^2 \;\Rightarrow\; \tilde{g}_{MS} = \frac{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}}, \tag{2}$$

where we have used conventional least squares to solve for the maximum SDR gain estimate $\tilde{g}_{MS}$.

A third intuitive solution is the matched power solution. This solution forces $\boldsymbol{x}$ and $\tilde{\boldsymbol{y}}$ to have the same power and it is also the geometric mean of the two previous solutions:

$$\tilde{g}_{MP} = \sqrt{\frac{\left| \boldsymbol{y} \right|^2}{\left| \boldsymbol{x} \right|^2}} = \sqrt{\tilde{g}_{MD}\, \tilde{g}_{MS}}\ . \tag{3}$$

In the decibel domain the geometric mean of (3) becomes an arithmetic mean:

$$\tilde{G}_{MP} = \frac{\tilde{G}_{MD} + \tilde{G}_{MS}}{2}, \quad \text{where } \tilde{G}_x = 20 \log_{10} \left(\left| \tilde{g}_x \right| \right). \tag{4}$$

Note that $\tilde{g}_{MD}$ and $\tilde{g}_{MS}$ will correctly detect a negative gain value (indicating a phase inversion in *S*) but $\tilde{g}_{MP}$ as defined in (3) will not. Thus we use an intuitive extension to redefine $\tilde{g}_{MP}$ as

$$\tilde{g}_{MP} = \text{sign}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y})\frac{|\boldsymbol{y}|}{|\boldsymbol{x}|}, \tag{5}$$

so that all three solutions will have the same sign.

3. ALGEBRAIC AND GEOMETRIC OBSERVATIONS

If we define the normalized input-output cross correlation (or input-output direction cosine) ρ as

$$\rho = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|}, \qquad -1 \le \rho \le 1, \tag{6}$$

then we can summarize the three solutions as

$$\begin{aligned} \tilde{g}_{MD} &= \frac{|\boldsymbol{y}|}{|\boldsymbol{x}|}\rho^{-1}, \\ \tilde{g}_{MP} &= \frac{|\boldsymbol{y}|}{|\boldsymbol{x}|}sign(\rho), \\ \tilde{g}_{MS} &= \frac{|\boldsymbol{y}|}{|\boldsymbol{x}|}\rho^{+1}. \end{aligned} \tag{7}$$

Equation (7) makes it clear that all three solutions have the same sign and that their magnitudes can be ordered:

$$|\tilde{g}_{MS}| \le |\tilde{g}_{MP}| \le |\tilde{g}_{MD}|. \tag{8}$$

In the limit as *S* becomes distortionless, $|\rho| \to 1$ and (7) makes it clear that the three solutions converge to a single (matched power) solution as expected.

Equation (7) also reveals that each of the three solutions is a special case of the more general solution

$$\tilde{g}(\alpha) = sign(\rho)\frac{|\boldsymbol{y}|}{|\boldsymbol{x}|}|\rho|^{\alpha}, \ -1 \le \alpha \le 1, \tag{9}$$

where α = -1, 0, and +1 correspond to the minimum distortion, matched power, and maximum SDR solutions respectively. Figure 2 provides an example of the geometric relationships among $\boldsymbol{x}$, $\boldsymbol{y}$, and the three possible gain-compensated outputs $\tilde{\boldsymbol{y}}_{MD}$, $\tilde{\boldsymbol{y}}_{MP}$, and $\tilde{\boldsymbol{y}}_{MS}$.

Note that these three solutions allow for arbitrary distortions in *S* but they also reproduce solutions that come from the minimization of additive input or output noises. The case of input noise leads to the minimum distortion solution. That is, if we solve for $\tilde{g}$ to satisfy $\boldsymbol{y} = \tilde{g}\left(\boldsymbol{x} + \boldsymbol{n}_{in}\right)$ while minimizing $\left|\boldsymbol{n}_{in}\right|^2$, we will arrive at $\tilde{g} = \tilde{g}_{MD}$. The case of output noise leads to the maximum SDR solution. That is, if we solve for $\tilde{g}$ to satisfy $\boldsymbol{y} = \tilde{g}\boldsymbol{x} + \boldsymbol{n}_{out}$ while minimizing $\left|\boldsymbol{n}_{out}\right|^2$, we will arrive at $\tilde{g} = \tilde{g}_{MS}$. The matched power solution $\tilde{g}_{MP}$ assumes that all output power is scaled input power and thus corresponds to the noise-free case $\boldsymbol{y} = \tilde{g}\boldsymbol{x}$.

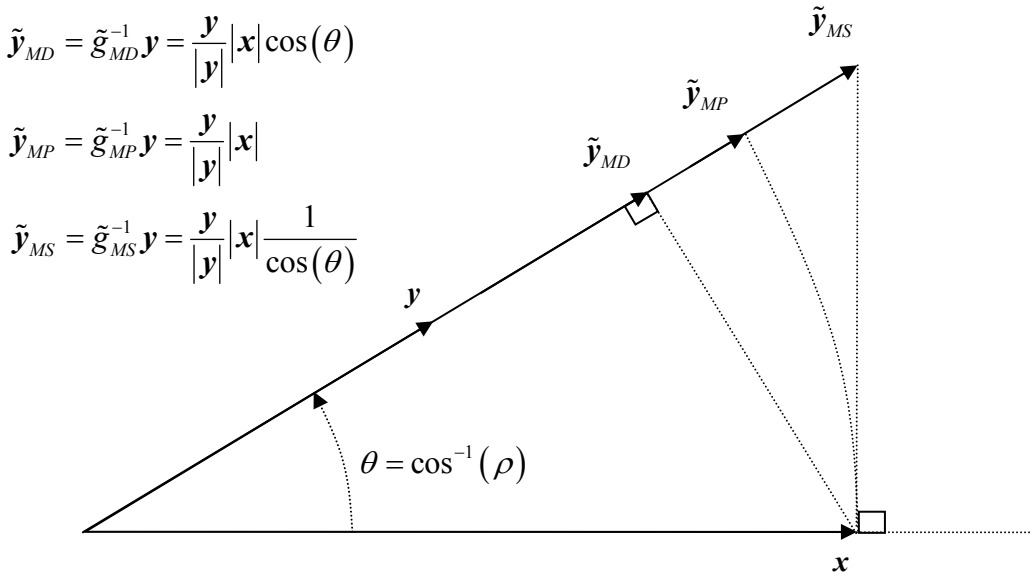

Figure 2. Example geometry of the three gain-compensation solutions.

4. EXAMPLE GAIN COMPENSATION RESULTS

In this section we present gain compensation examples for speech and video systems. In the speech example the input $\boldsymbol{x}$ is a ten-second speech signal and the system *S* is a 5.3 kbps speech codec conforming to ITU-T Recommendation G.723.1 [10]. In the video example the input $\boldsymbol{x}$ is one frame of a video chrominance signal and the system *S* consists of several play and record cycles of an analog video tape recorder. Figure 3 and Figure 4 show the SDR and distortion as a function of $\tilde{G}$ for the speech and video examples respectively. The calculated values of $\tilde{G}_{MS}$,

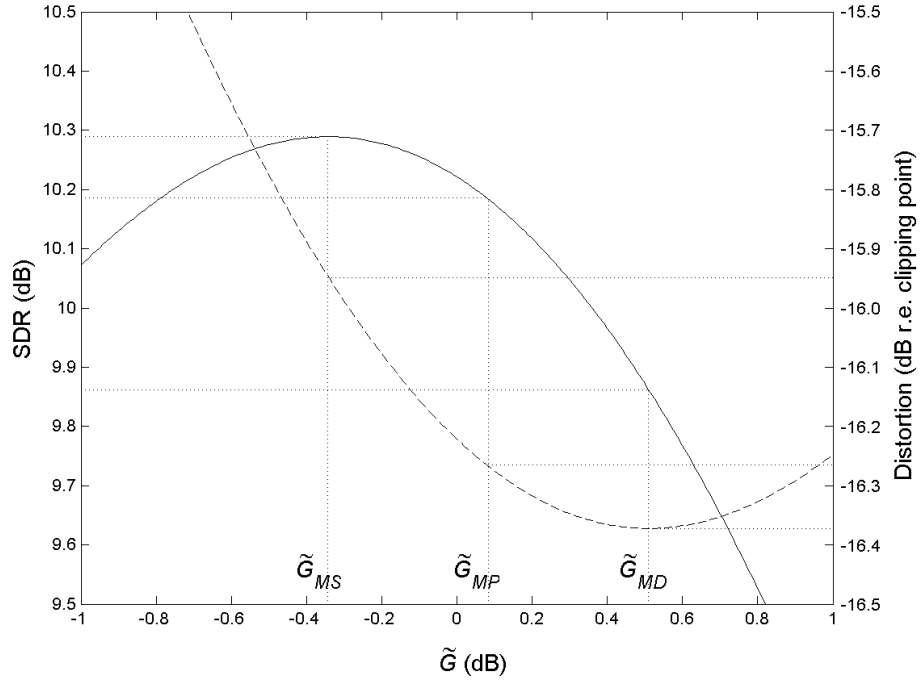

Figure 3. Example gain-compensated SDR (solid line) and gain-compensated distortion (dashed line) vs. estimated gain ($\tilde{G}$) for a speech codec.

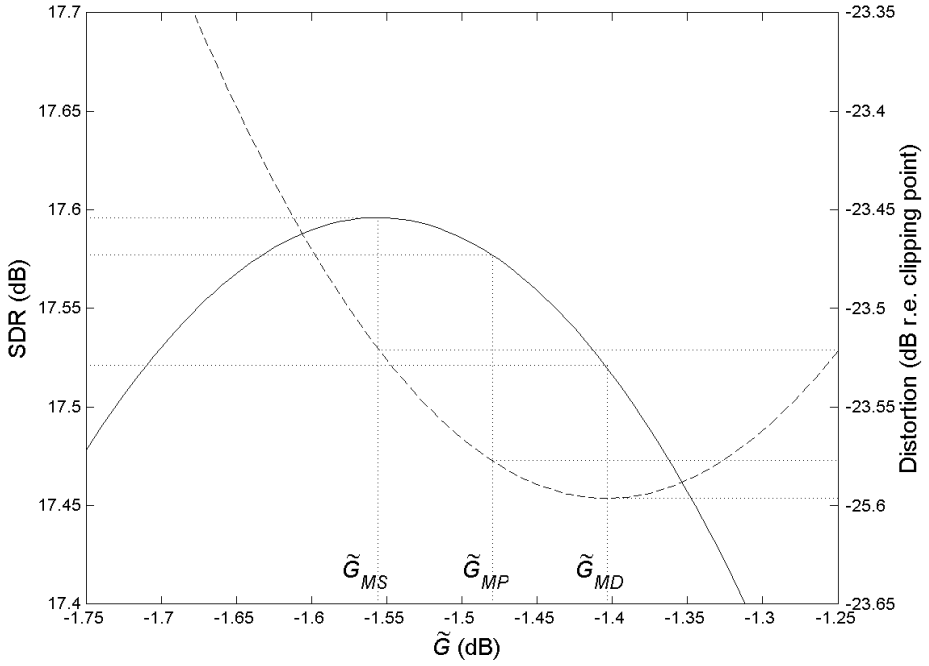

Figure 4. Example gain-compensated SDR (solid line) and gain-compensated distortion (dashed line) vs. estimated gain ($\tilde{G}$) for an analog video tape recorder.

$\tilde{G}_{MP}$, and $\tilde{G}_{MD}$ are marked. As expected, $\tilde{G}_{MS}$ corresponds to the SDR maximum, $\tilde{G}_{MD}$ corresponds to the distortion minimum, and $\tilde{G}_{MP}$ is midway between $\tilde{G}_{MS}$ and $\tilde{G}_{MD}$. Note that in the speech example, the values of the three gain compensation solutions span a range of about 0.8 dB, and the resulting SDR and distortion values span a range of about 0.4 dB. For the video example these ranges are smaller. In general, these ranges will depend on both the input $\boldsymbol{x}$ and the system S.

5. DISCUSSION

We have established that there are multiple well-motivated solutions to the gain compensation problem described in Figure 1, and that these solutions can differ significantly in real applications. We might ask which solution should be used in practice but of course there is no single answer. Each solution does exactly what its name says it does, so the question of which solution to use becomes the question of which attribute is most relevant in a given context. Do we wish to minimize distortion, match power, or maximize SDR?

We could also ask the question somewhat differently in terms of $\tilde{g}$. Loosely speaking, we could ask if $\tilde{g}^{-1}$ should describe how to scale the output to make it "as close as possible" to the input (leading to the minimum distortion solution), or if $\tilde{g}$ should describe the output-to-input power ratio (leading to the matched power solution), or if $\tilde{g}$ should describe how to scale the input to make it "as close as possible" to the output (leading to the maximum SDR solution)?

We could ask the question in a third way, using even more informal language. Should $\tilde{g}^{-1}$ describe the fraction of the output that "matches" the input (leading to the minimum distortion solution where we project the input vector onto the output vector), or should $\tilde{g}$ describe the output-to-input power ratio (leading to the matched power solution), or should $\tilde{g}$ describe the fraction of the input that "matches" the output (leading to the maximum SDR solution where we project the output vector onto the input vector)?

As noted above, the nature of human auditory and visual perception may make SDR more relevant than distortion in audio and video systems. This might lead one towards a maximum SDR solution for the gain compensation problem in the audio and video system context. We note further that "distortion" as we have used it here is waveform distortion. In perception-based audio codecs, waveforms may be severely distorted while the perceived audio signal is minimally distorted. For these and similar systems, the matched power solution may be more appropriate than the minimum waveform distortion solution or the maximum waveform SDR solution. For the audio or video quality estimation application, it may be even more appropriate to compensate for system gain by the matching of some estimates of perceived loudness or contrast.

A variety of gain compensation choices can be found in the objective audio and video quality estimation literature. The audio quality estimation algorithm in [6] includes a level adaptation stage that effectively calculates $0 < \tilde{g}_{MD}$ for frequency-domain excitation patterns. When $\tilde{g}_{MD} \le 1$ the input is scaled by $\tilde{g}_{MD}$, otherwise the output is scaled by $\tilde{g}_{MD}^{-1}$. The speech quality estimation algorithms given in [2] and [3] effectively perform matched power gain compensation, but they

apply different scale factors to both the input and output signals to bring them to a common fixed level (e.g. unit variance). The speech quality estimation algorithm in [4] performs matched power gain compensation and applies a scale factor to the output only. The video quality estimation algorithm in [8] uses an iterative algorithm to find a robust estimate of $\tilde{g}_{MS}$. This algorithm uses weighted least-squares and places smaller weights on samples that have greater distortion so that they will not unduly influence the gain estimate. The resulting scale factor is applied to the output only.

Finally we mention system gain compensation via input scaling as shown in Figure 5. This is a somewhat less intuitive but still reasonable alternative to the output scaling approach given in Figure 1. If the input scaling approach is adopted, then the minimum distortion and maximum SDR solutions will be identical to each other and will be given by (2). The matched power solution will be unchanged and thus will be given by (5).

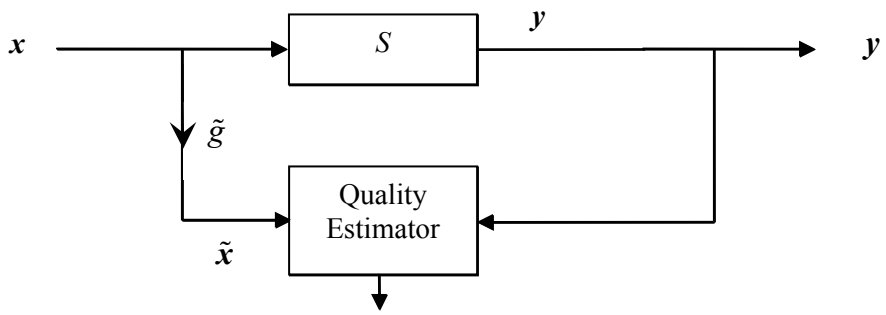

Figure 5. Block diagram for gain-compensated quality estimation with input scaling.

6. CONCLUSIONS

There are three mathematically-motivated solutions to the gain compensation problem described in Figure 1. The solution used should reflect a conscious choice based on the system under consideration. Depending on which attribute is most relevant, we may choose to seek a gain compensation that minimizes distortion, matches power, or maximizes SDR. We have derived these three solutions, described the algebraic and geometric relationships between them, and provided a generalized result that subsumes all three. We have demonstrated that these solutions can differ significantly in real applications (e.g. 0.8 dB for the G.723.1 speech coder) and have reported some of the gain compensation choices found in the objective audio and video quality estimation literature.

7. REFERENCES

[1] A.W. Rix, J.G. Beerends, M.P. Hollier, and A.P. Hekstra, "Perceptual evaluation of speech quality (PESQ) – A new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing 2001*, Salt Lake City, May 2001, vol. 2, pp. 749-752.

[2] 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," Geneva, 2001.

[3] S.D. Voran, "Objective estimation of perceived speech quality, part II: Evaluation of the measuring normalizing block technique," *IEEE Transactions on Speech and Audio Processing,* vol. 7, no. 4, pp. 383-390, Jul. 1999.

[4] ITU-T Recommendation P.861, "Objective quality measurement of telephone-band (300-3400 Hz) speech codecs," Geneva, 1996.

[5] T. Thiede, W.C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J.G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ – The ITU standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, Jan./Feb. 2000.

[6] ITU-R Recommendation BS.1387, "Method for objective measurements of perceived audio quality," Geneva, 2001.

[7] S. Wolf, "Measuring the end-to-end performance of digital video systems," *IEEE Transactions on Broadcasting*, vol. 43, no. 3, pp. 320-328, Sep. 1997.

[8] S. Wolf and M. Pinson, "Video quality measurement techniques," NTIA Report 02-392, Jun. 2002 (available at http://www.its.bldrdoc.gov/n3/video/documents.htm).

[9] C.J. van den Branden Lambrecht, D.M. Costantini, G.L. Sicuranza, and M. Kunt, "Quality assessment of motion rendition in video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 5, pp. 766-782, Aug. 1999.

[10] ITU-T Recommendation G.723.1, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s," Geneva, 1996.