IETF RMCAT WG: Video Quality Metrics Discussion for Evaluation Criteria

V. S. Somayazulu
(v.srinivasa.somayazulu@intel.com)
Hassnaa Moustafa
(hassnaa.moustafa@intel.com)

Problem Statement

- ***Video Quality Perception*** is critical for end-user QoE
  - Topic has been discussed in WG, but no conclusion yet on including this in the requirements or the evaluation metrics for the congestion control algorithms
  - Network/transport related metrics are currently present in the evaluation criteria for Rmcat congestion control algorithms, but do not fully capture the impact on video QoE.

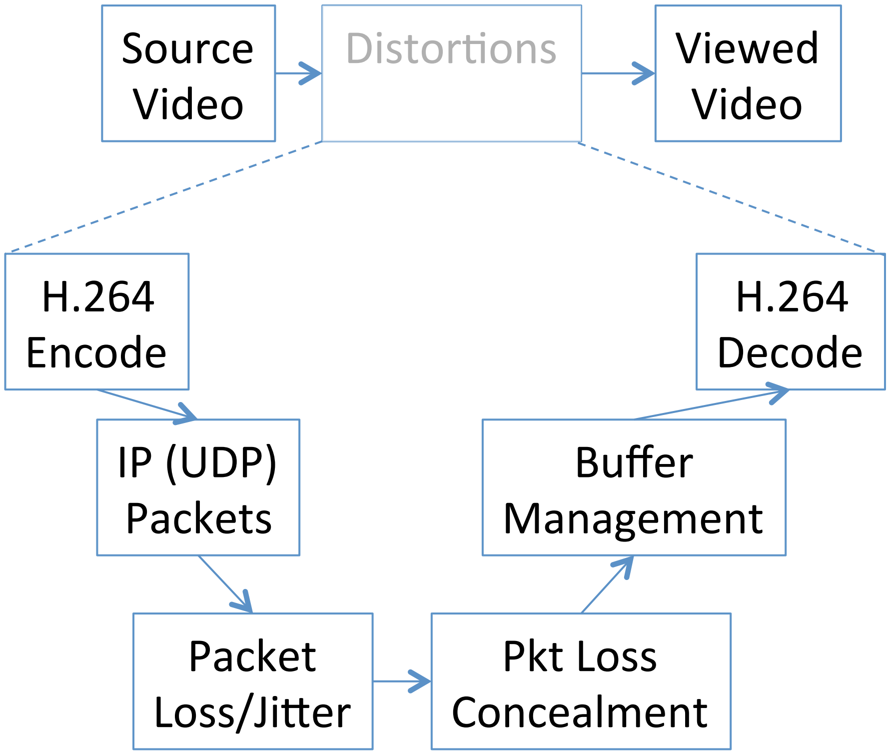
Purpose of this presentation:
1. Provide background on video quality metrics that can help quantify impact of congestion control on the video QoE
2. Stimulate a discussion on adoption of appropriate metrics in evaluation criteria

(High Level) Summary of Evaluation Scenario Discussions in RMCAT WG

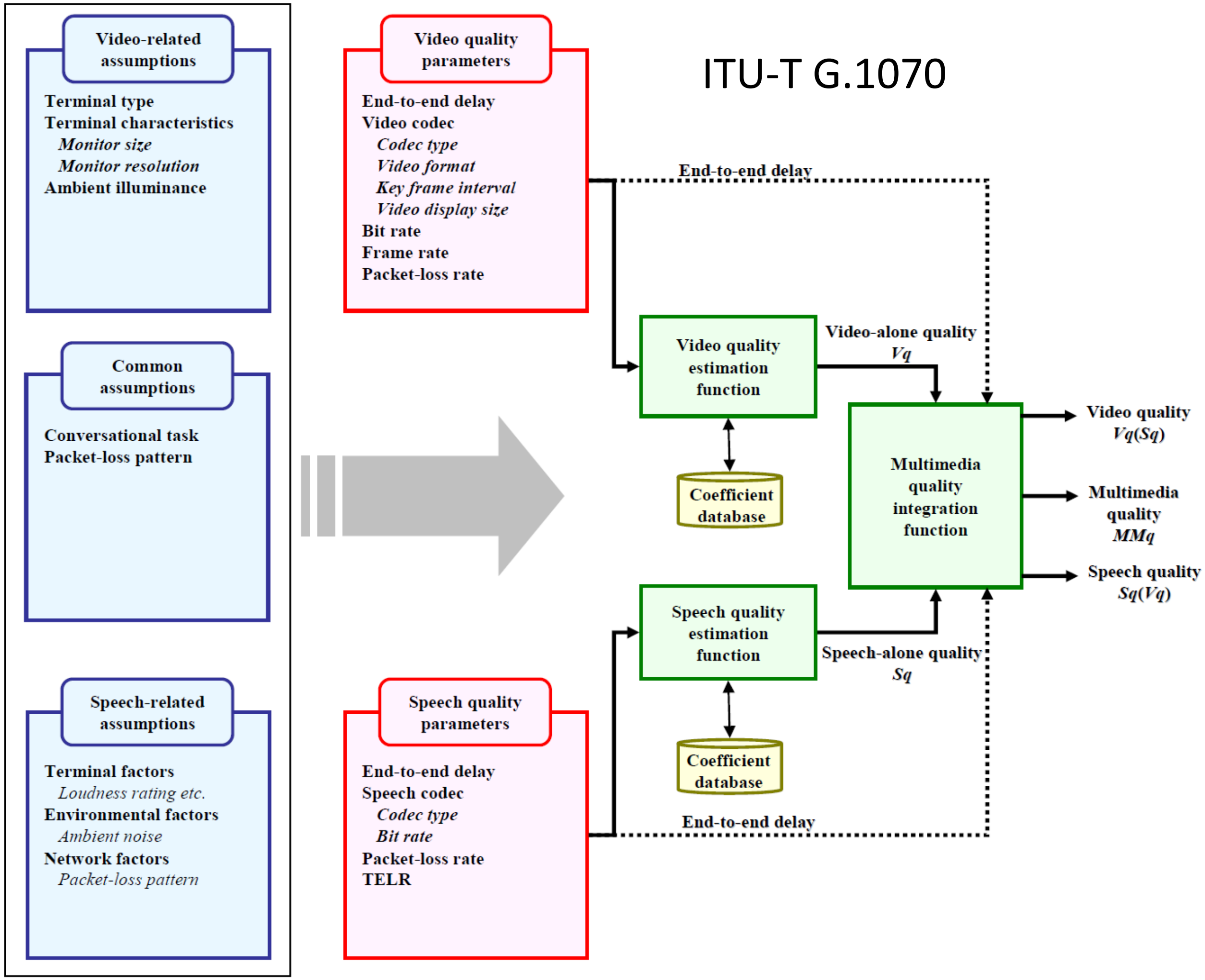
- Current (evolving) direction seems to be - Evaluate overall performance, i.e., Video/Audio+ Network Congestion control algorithm
- Couple of ways to include video characteristics:
  1. Use a traffic model that captures statistics of the video source+rate control+shaping
  2. Possibly use uncompressed video sequences with a live encoder (+ prescribed settings) and congestion control algo.
- With either approach, we should define a means of capturing impact on the overall video QoE.

Objective Metrics – Big Picture View

- Background
  - Objective metrics developed to mimic human perception of video, e.g., "look at the received video, and judge how good it looks"
  - Distortions between source and video considered a "black box" and did not affect metric design
    - Good for unbiased evaluation of encoders, etc.
  - Non-reference / Reduced reference a difficult problem in this generic case
- Full-reference vs. non-reference VQM
  - Full reference: compare the measured video with the original uncompressed video
    - PSNR, MS-SSIM, PEVQ
  - Non-reference: analyze the video without a comparison
    - E.g. P1202, etc.
    - ITU-T G.1070



Video Conferencing QoE – ITU-T G.1070



Recommendation ITU-T G.1070 proposes an algorithm that estimates videophone quality for quality of experience (QoE)/quality of service (QoS) planners.

ITU-T G.1070 Video Quality Metric

- Video quality is calculated as: $$V_q = 1 + I_{coding} \exp\left(-\frac{Ppl_V}{D_{PplV}}\right)$$
  - where I_{coding} represents basic video quality as a function of video bitrate and frame rate
  - D_{PplV} represents degree of video quality robustness due to packet loss and *PplV* is the packet loss rate in %
  - These quantities are calculated using a set of fixed parameters dependent on codec type, video format, key frame interval, and video display size
    - G.1070 provides provisional values for H.264, VGA format, 1 second key frame interval and 9.2 inch display, coded bit rates between 400 kbps – 2 Mbps, packet loss rates < 5% and frame rates from 5-25 fps.
    - Parameter values for modeling any other set of conditions would need to be derived from video quality evaluation
- Further enhancements to the basic model for different codecs, display formats, content dependency, etc. have been explored, e.g.:
  - [Joskowicz, 2009] Joskowicz, J. et al "Enhancements to the Opinion Model for Video-Telephony Applications", *Proc. 5th Latin American Networking Conference,* pp. 87-94.

Considerations on use of ITU-T G.1070

- Use evaluation scenario(s) to run simulation of target congestion avoidance algorithm with a given test video sequence (trace file)
  - Collect trace of packet arrival times, packet losses, at sender and receiver
  - Segment data into short intervals of time (e.g. 5 seconds?)
  - For each segment *i*:
    - Calculate video bitrate sent, assume fixed frame rate
    - Calculate average packet loss rate at receiver
    - Calculate per-segment video quality $V_{q,i}$
  - Calculate over the entire video sequence:
    - Mean and variance of the set of $\{V_{q,i}\}$
    - Higher mean and lower variance → better overall video quality

Discussion

- Propose to add video quality evaluation metrics for consideration by Rmcat WG. Desirable features of the metrics should include:
  - Good correlation with subjective video quality perception
  - Combine different parameters to provide an integrated look at video QoE impact
  - Relatively simple to calculate based on data from network simulations
  - Ideally, based on published standards
- Subjective quality evaluation is hard to organize and execute, especially in a contribution evaluation phase
- Common objective evaluation metrics of video quality
  - Easier to use in proposal evaluations,
  - Full-reference & Non-reference: latter may be more suited for RMCAT evaluation

- ITU-T Rec.G.1070 based NR VQ metrics designed for video conferencing applications
  - Additionally, this doesn't require compressed bitstream inspection, etc.: algorithm inputs are high level, e.g. throughput, packet loss rate, etc.

- **Soliciting feedback from group on defining consideration for video quality metrics as part of evaluation criteria for congestion control algorithms? → Current phase**

- **For further consideration: Could VQ be exploited by congestion control algorithms?**
  - **Potential for incorporating video quality information into RTCP XRBLOCK reports?**

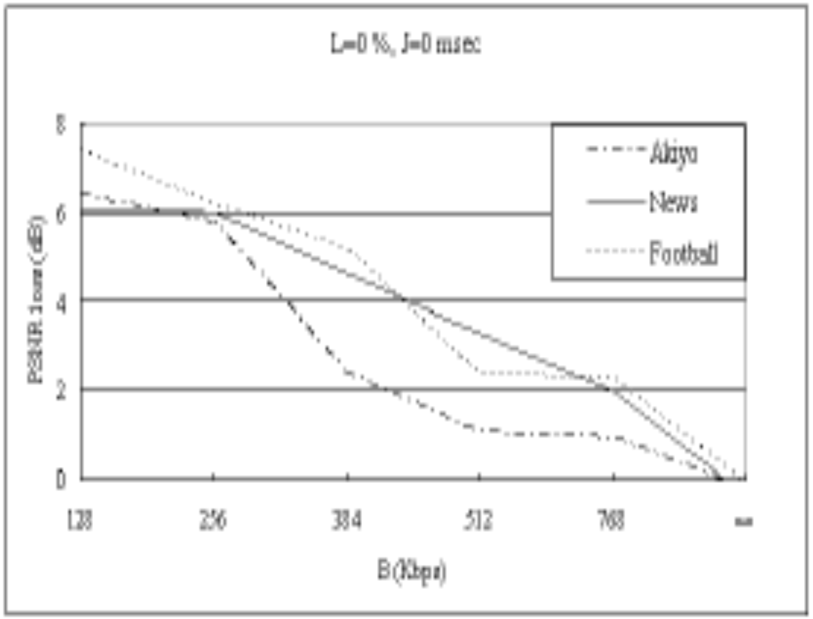
  → (Later phase)

Annex

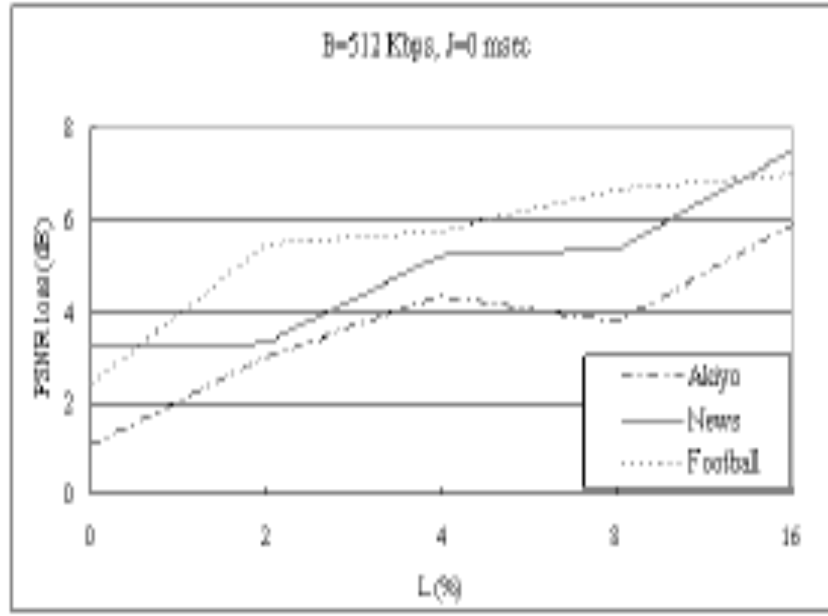
Background

- Rmcat WG is dealing with congestion control for Internet data considering interactive point-to-point real-time multimedia services over RTP
- Requirements for congestion control algorithms are defined considering
  - Low delay
  - Semi-Reliable data delivery
  - Fairness to other flows
  - Adaptation to network conditions
- Metrics for congestion control are defined to be
  - Delay, throughput, minimizing transmission rates oscillations, reactivity to transient events and packet losses and discards
- Evaluation Criteria for congestion control algorithms have been defined considering
  - Avoiding Congestion Collapse
  - Stability
  - Media Traffic
  - Startup Behavior
  - Diverse Environments
  - Varying Path Characteristics
  - Reacting to Transient Events or Interruptions
  - Fairness with Similar Cross Traffic
  - Impact on Cross Traffic
  - Extensions to RTP/RTCP

Video Quality Variation w/ Network Conditions Variation



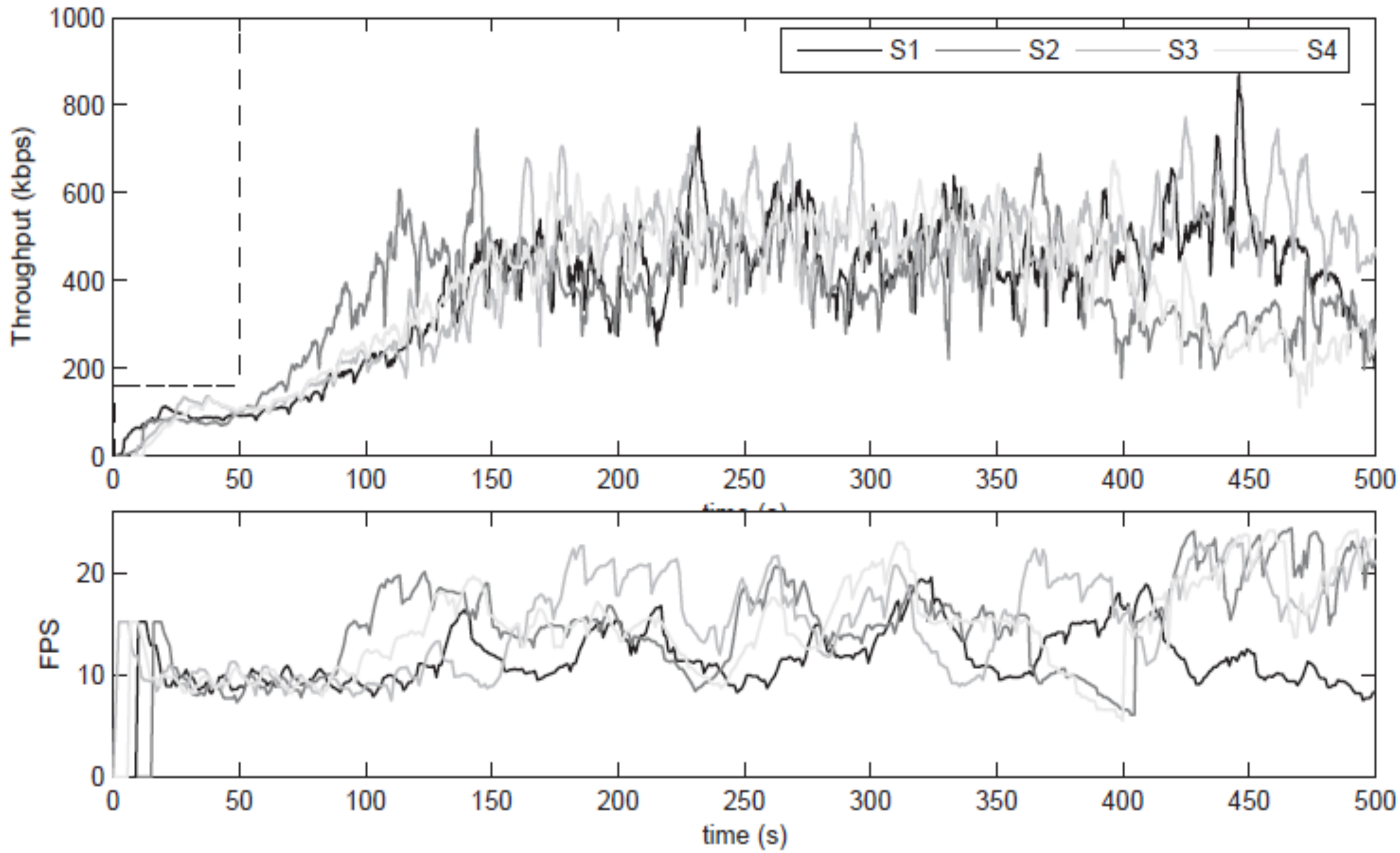
PSNRloss for different network bandwidth limitations



PSNRloss for different packet loss rate

Source: R-G Cheng et al., "Measurement and Analysis of Skype Video Traffic," APWCS 2008.

Impact of Throughput Variation on Video Conferencing Applications



Source: L. De Cicco et al., "Skype Video Responsiveness to Bandwidth Variation," ACM NOSSDAV 2008.

Video Quality Evaluation: Introduction

- Quality of Experience
  - The overall experience the consumer has when accessing and using provided video services
- Quantifying Video Quality
  - Mean Opinion Score (MOS)
    - Subjectively done: recruit a group of people to watch a set of video clips and give a numeric score to each clip
    - Automatically done: design algorithms to estimate a MOS based on characteristics of media stream, network, device, etc.
- Video Quality Issues
  - Video creation, video encoding/transcoding, video transmission, video display

Video Quality Issues

- Video blockiness (encoding)
- Video blurriness (capturing/encoding)
- Video losses (transmission)
- Video jerkiness (transmission/encoding/display)
- Video freezing/rebuffering (transmission)
- A/V sync problem (transmission/encoding)

Blockiness vs. blurriness

PSNR (Peak Signal-to-Noise Ratio)

- Most commonly used metric to measure the quality of reconstruction of lossy compression codecs

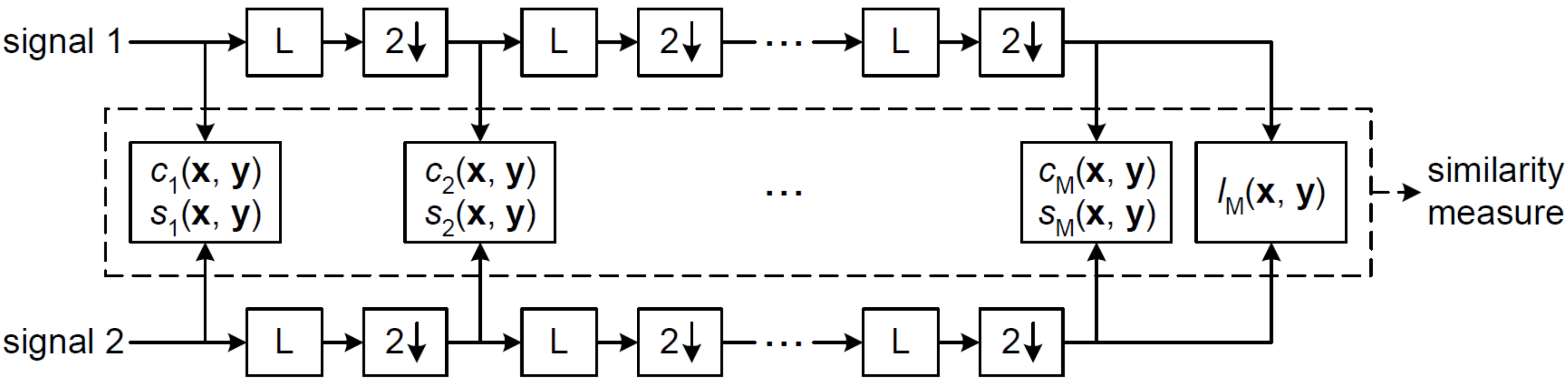
$$PSNR = 10 \cdot \log_{10} \left(255^2 / MSE \right)$$

  - Typically values between 30~50 dB, higher is better

MS-SSIM

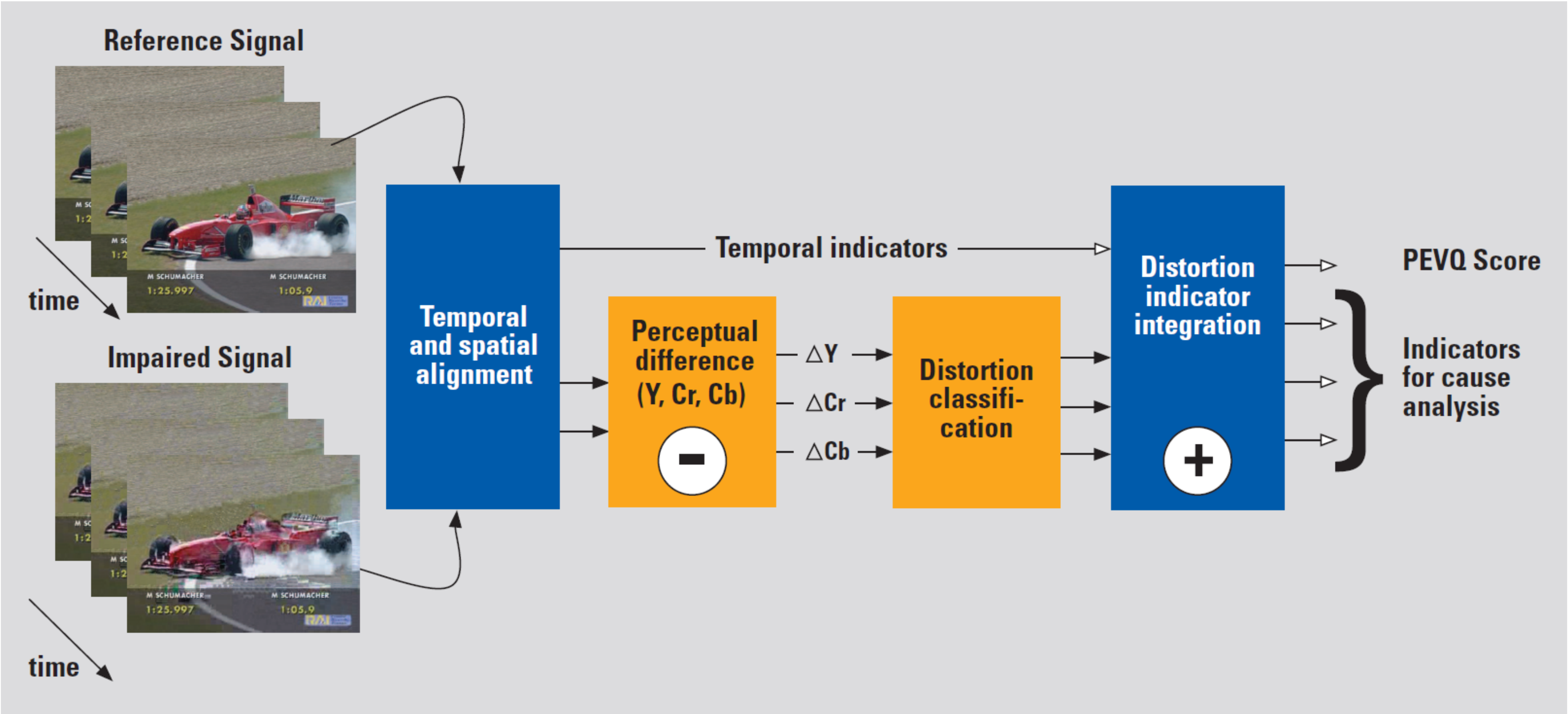
$$MSSSIM(\boldsymbol{x},\boldsymbol{y}) = [l_M(\boldsymbol{x},\boldsymbol{y})]^{\alpha_M} \prod_{j=1}^{M} [c_j(\boldsymbol{x},\boldsymbol{y})]^{\beta_j} [s_j(\boldsymbol{x},\boldsymbol{y})]^{\gamma_j}$$

Luminance *l(x, y)* Contrast *c(x,y)* Structure s(x,y)



Typical values between 0.8~1, higher is better

PEVQ

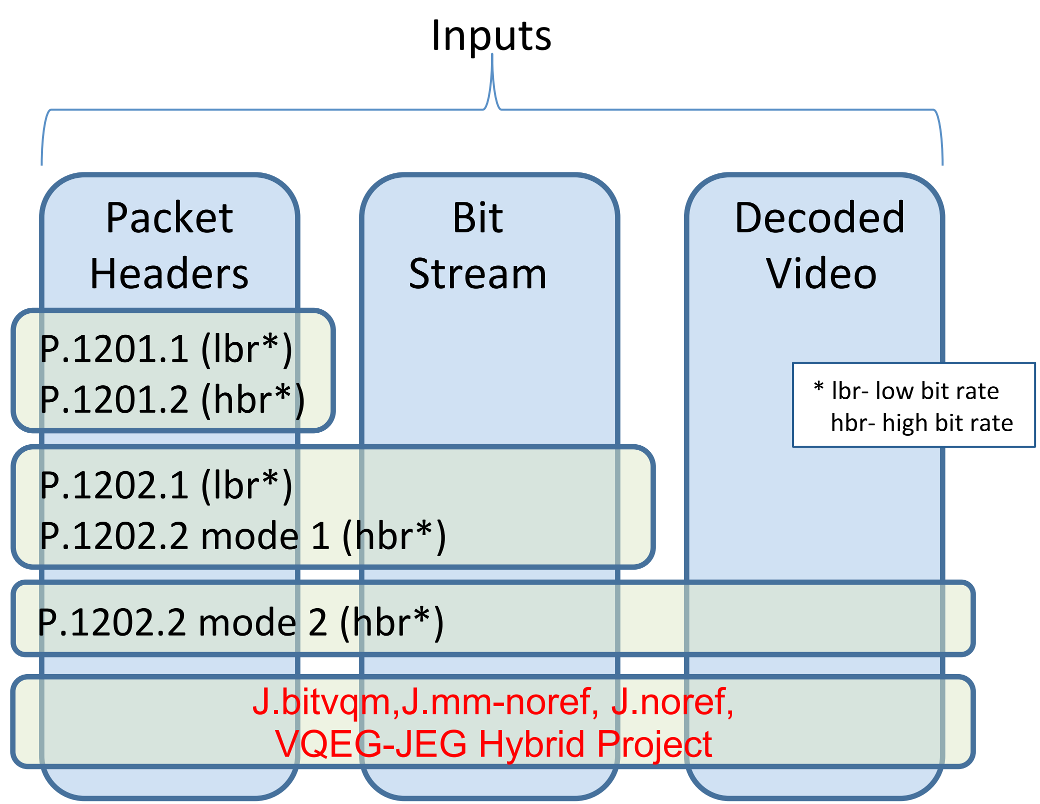


Output PEVQ MOS ranging from 1 (bad)~5 (excellent)

Distortion Indicators: Delay, Brightness, Contrast, PSNR, Jerkiness, Blur, Blockiness, Frame skips and freezes, Temporal Activity and spatial complexity

ITU-T SG12 No-Reference Objective Standards

- Consented or Approved documents shown in **black**
- Targeted Applications: IPTV services; non-adaptive streaming; non-interactive
- P.1201.x – multimedia QoE; P.1202.x – video QoE
- Targeted protocols: non-HTTP (RTP, TS-on UDP, etc.)



P.120x.x Video Impairment Model

- P.1202 Approach
  - Design a metric for evaluating quality in specific instance; i.e., IPTV services
  - This *significantly* constrains the problem, and should make the solution much more feasible
    - Known video encoder; encoder output available
    - Known channel; impairment pattern available
- Four main video distortions accounted for
  - Compression Artifacts
    - Due to lossy encoding
  - Slicing Artifacts
    - Due to Packet Loss Concealment (PLC) of lost packets
  - Freezing Artifacts
    - Due to PLC replacing erroneous frames with last good frame ("freezing with skipping")
  - Rebuffering Artifacts
    - Due to PLC repeating a frame until frame reception recommences ("freezing without skipping / spinning wheel")

P.1202 (ex. P.NBAMS)

- Two application areas :
- P.1202.1,"lower resolution mode”:
  - Same as P1201.1:
    - i.e., QCIF-QVGA-HVGA, mostly for mobile TV and Streaming
- P.1202.2, "higher resolution mode”:
  - Linear broadcast TV & Video on-demand:
    - Still under Study

P.1202 (ex. P.NBAMS)

- Packet Headers and bitstream input only
- Not intended for codec evaluation
- Not intended for streams with significant rate adaptation
- Video Pearson correlation of 0.918 for P.1202.1 (982 samples)

Validated Test Factors		
Recommendation	P.1202.1	P.1202.2
Audio BR	NA	NA
Video BR	0.05 – 6 Mbps	Under study
Packet loss	✓	Under study
Re-buffering	✓	
Video Resolution	HVGA, QVGA, QCIF	Under study
Video encoding	H.264/AVC baseline	Under study
FR's and key frame rates	Frame rate 5-30 Hz GOP lengths 2-10s	Under study
Protocol	UDP-based	Under study
Protocols not tested*	TCP-based	Under study

* (can be used, but may not be reliable)