Tutorial A

Objective Measurement of User-Perceived Audio and Video Quality

Objective Measurement of User-Perceived Audio and Video Quality

Lectures:
Measurement of Audio Quality, Steve Voran, (≈30 min)
Measurement of Video Quality, Steve Wolf , (≈30 min)

Demonstration:
Real-Time, In-Service Video Quality Measurement, Margaret Pinson , (≈20 min)

All Presenters are with:

Institute for Telecommunication Sciences, Boulder Colorado
www.its.bldrdoc.gov/audio www.its.bldrdoc.gov/n3/video
{svoran, swolf, mpinson} @its.bldrdoc.gov

Combined A/V (MM) Quality

- Perceived overall A/V quality depends on
  - Audio quality
  - Video quality
  - A/V synchronization
  - Importance audio and video in application

Measurement of Audio Quality

- Motivation
- Subjective Measurement
  - Techniques
  - Issues
  - Open Questions
- Objective Estimation
  - Basic Concepts
  - Perception-based Approach
  - Standardized Algorithms
  - Moving Forward

One Major Motivation Telecommunications: Five-Way Trade-Off

- Complexity vs. delay vs. bit rate vs. robustness vs. speech quality
- The first four are fairly easily defined and measured
- Speech quality is more challenging
  - Constrained optimization of coders and/or channels
  - System monitoring and maintenance alerts
  - Picking "best" designs for standardization
  - Equipment evaluation for purchase

Demonstration: Speech Coding Algorithms

- All have nominal 3 dB bandwidth of 300-3400 Hz
- 4 Talkers: female, male, female, male
- Clear channel (no bit errors, lost packets or frames)

Demonstration: Speech Coding Algorithms

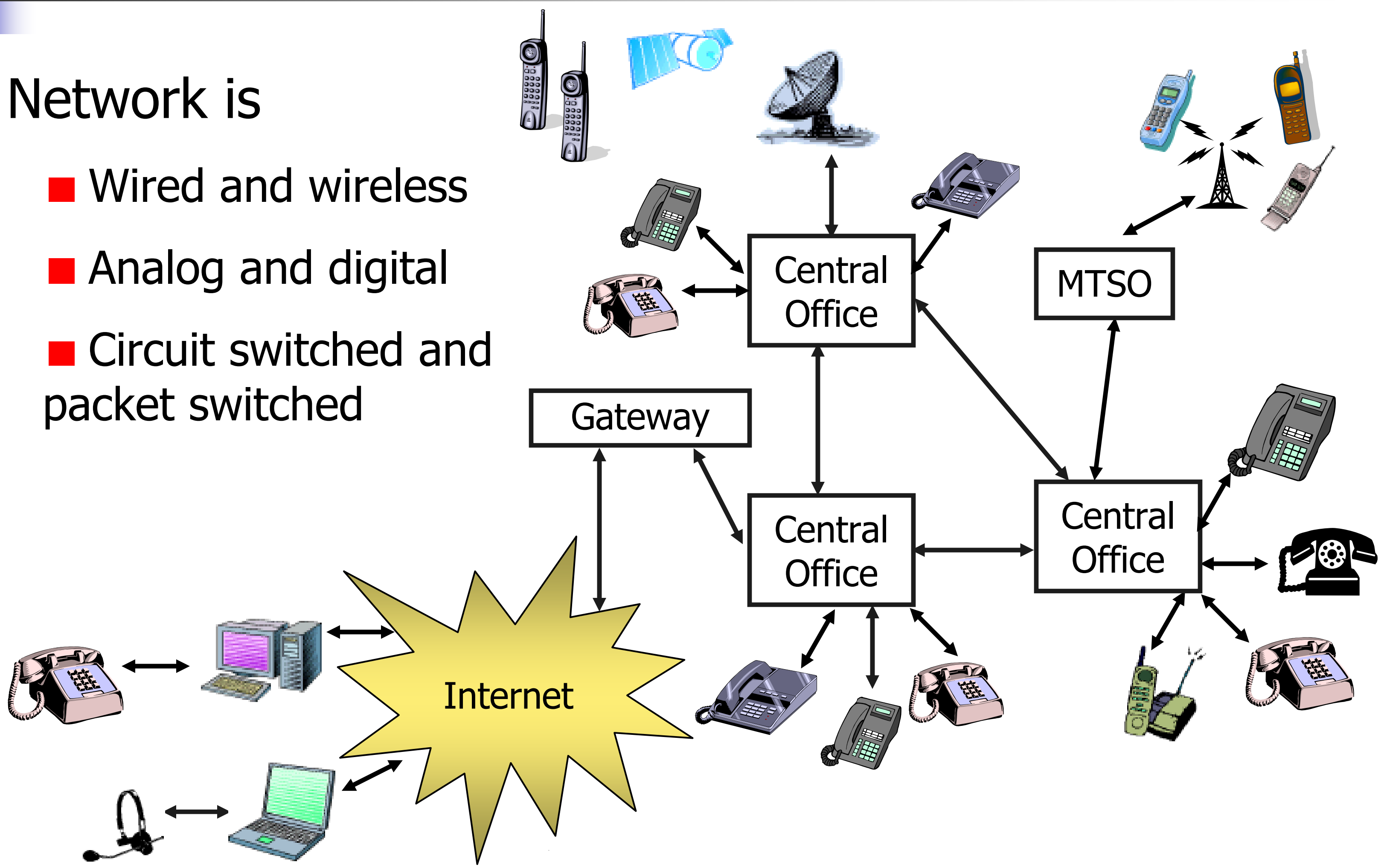
- Original Speech
- G.711 64 kbps PCM
- G.726 32 kbps ADPCM
- G.728 16 kbps LD-CELP
- GSM 13 kbps RPE-LTP
- G.729 8 kbps CS-ACELP
- G.723.1 5.3 kbps ACELP
- Fed. Standard 2.4 kbps MELP
- FS-1015 2.4 kbps LPC10e

Quality Assessment Challenges

- Signal dependent distortions
  - Male vs. female
  - Speech level
  - Language
  - Tones vs. speech
- Variety of distortion types
  - Buzz, robotic
  - Warble, shimmer
  - Muffled, flat
- Listener expectations
  - Fixed vs. wireless
  - Cost

Quality Assessment Challenges: Heterogeneous Network Paths

- Network is
  - Wired and wireless
  - Analog and digital
  - Circuit switched and packet switched



Quality Assessment Challenges

- Tandems, mixed tandems
- Channel issues:
  - Bit error rate, packet loss rate
  - Temporal distribution of errors
  - Delay variation
- Handset types and hands free
- Background noise – demo follows

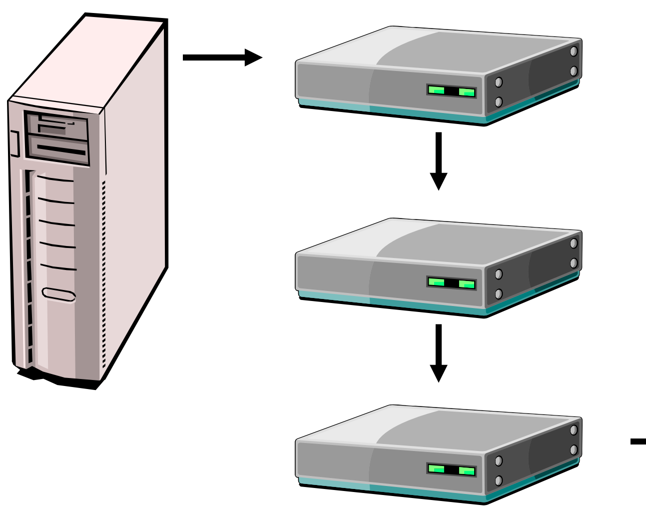
Background Noise Demo

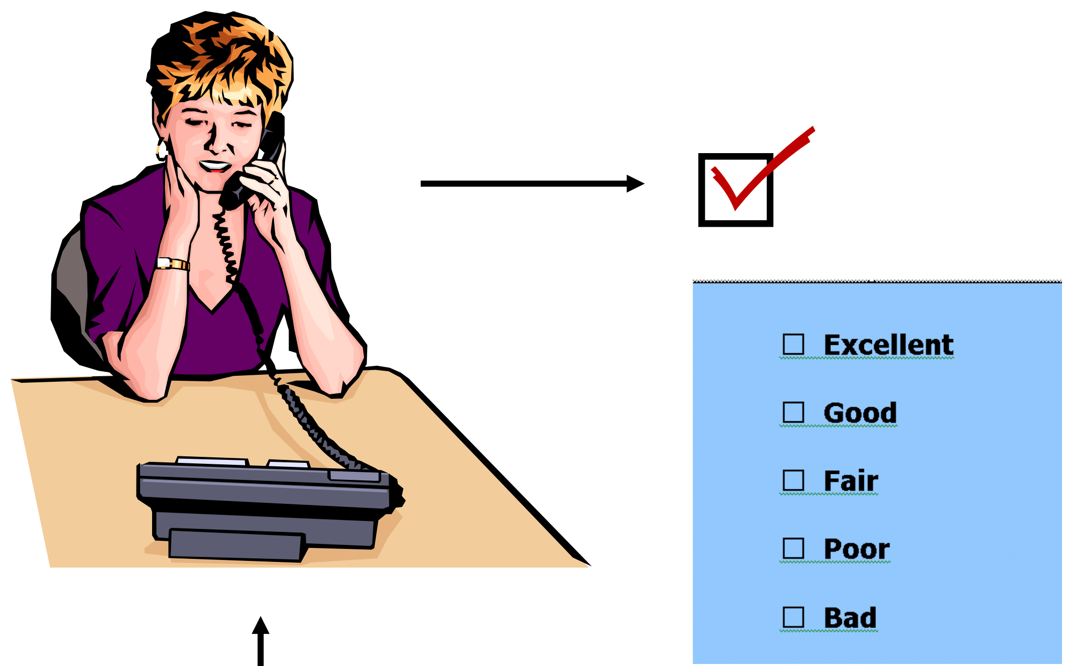
Speech Coder 4800 bits/second

- on airplane
  - original after speech coder
- in laundry room
  - original after speech coder
- on the boat with the radio
  - original after speech coder

Subjective Measurement: Have People Listen and Respond

Example: Absolute Category Rating Test





Subjective Measurement

- Advantages:
  - If done carefully, results are highly relevant (standards exist)
  - Can select relevant population of listeners
  - Can educate them about application
  - Can ask the most relevant question(s)

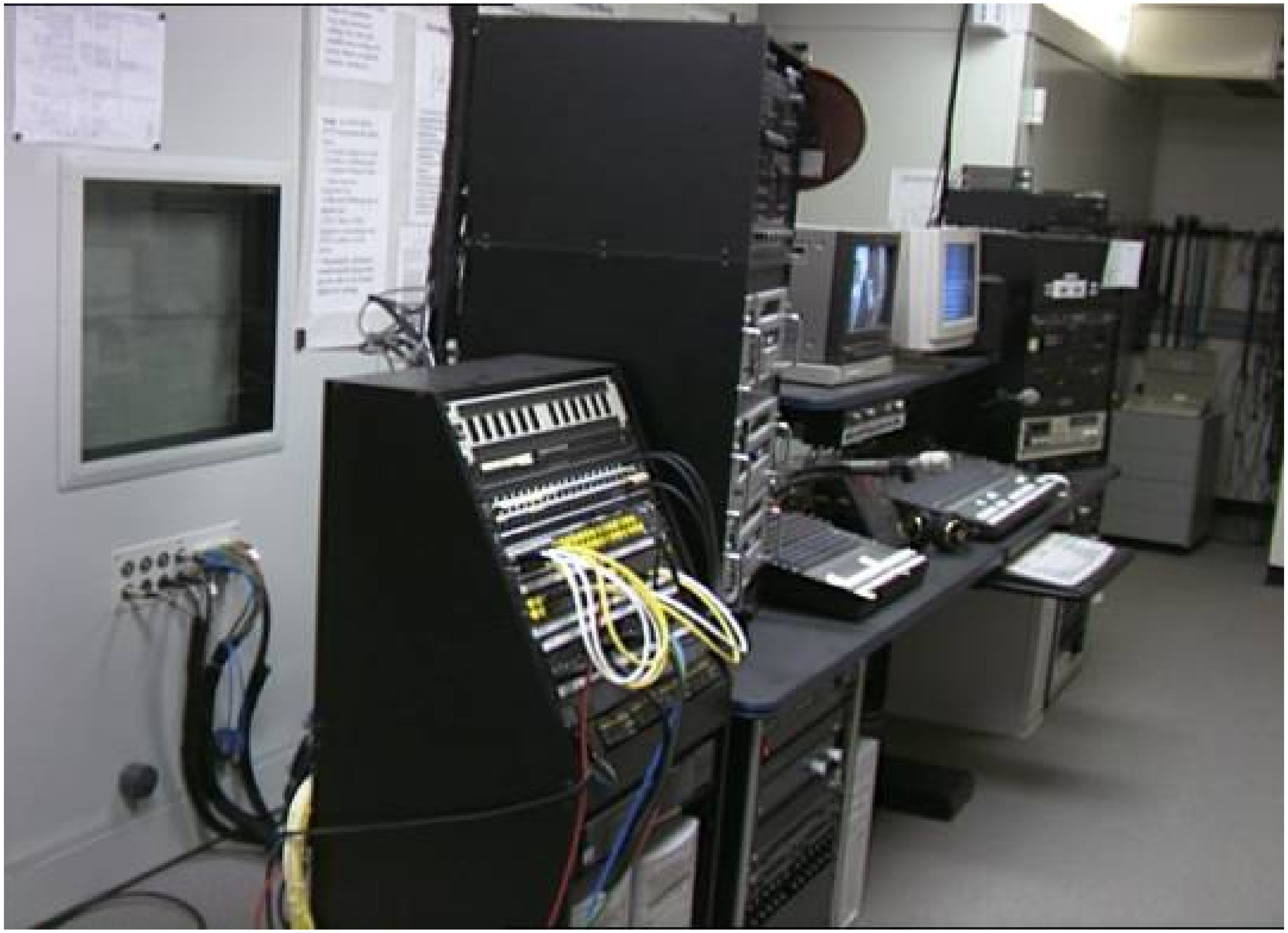
Subjective Measurement

- Disadvantages:
  - Expensive and time consuming
  - Controlled acoustic environment required
  - Transparent playback equipment required
  - Inherent spread of opinions - depending on confidence intervals required, 20 to 60 listeners may be required
  - Each listener may spend 30 to 120 minutes
  - Results not absolute, reference conditions required

ITS Subjective Measurement Capabilities

- ITS has two acoustically isolated rooms and associated laboratory equipment (pictures follow)
- Loudspeaker, headphone, or handset playback is available
- ITS has conducted numerous listening and conversation tests in conformance with applicable ITU Recommendations

Control Station for Subjective Testing Lab

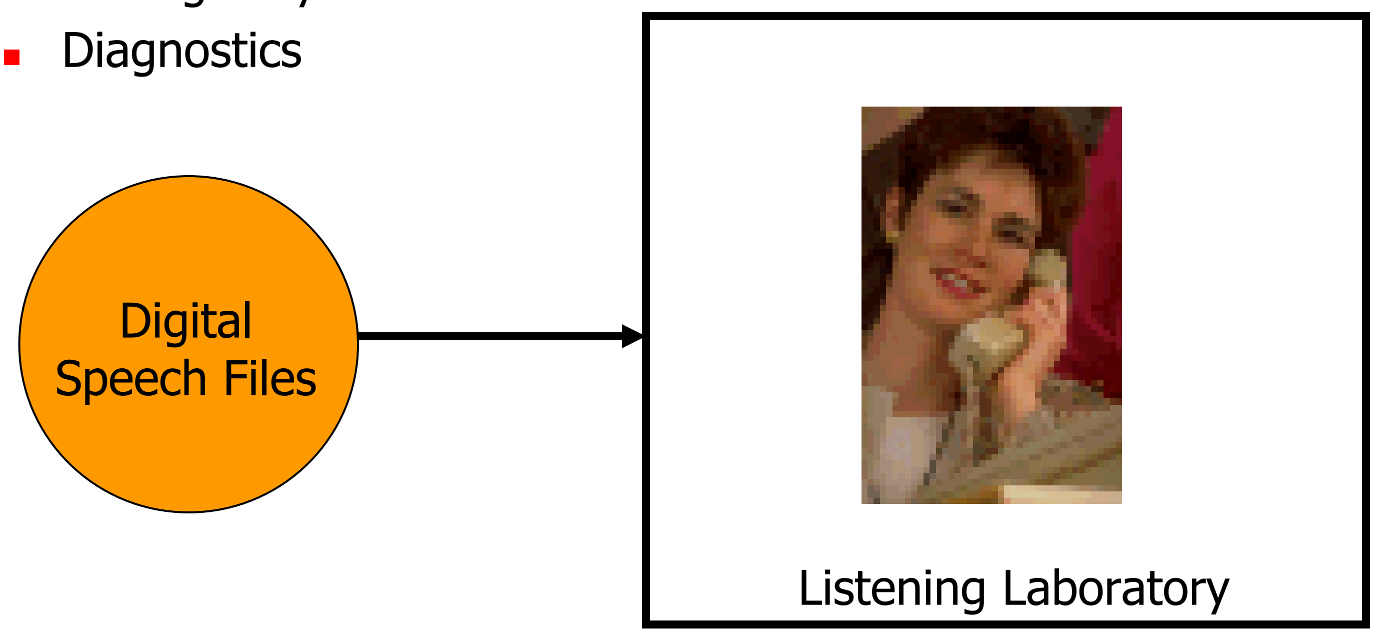


Inside a Subjective Testing Room

Subjective Listening Tests

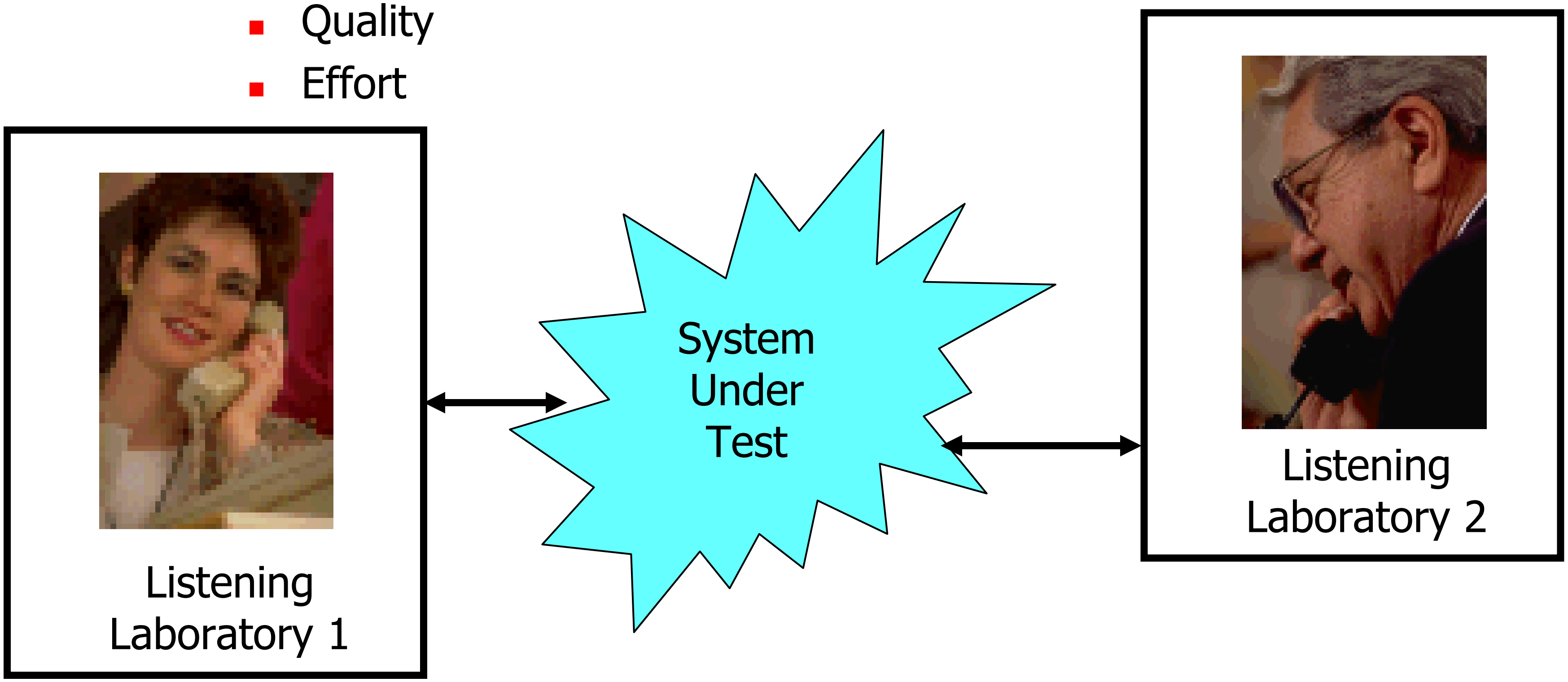
Subjects hear and score recordings

- Quality
- Effort
- Intelligibility
- Diagnostics



Subjective Conversation Tests

- Subjects use system under test for actual conversation
- Subjects score system under test
  - Quality
  - Effort



Considerations in Subjective Testing

- Listening instrument, expectations
- Background noise
- Selection of listeners: demographics, hearing acuity, prior knowledge, expectations
- Listener fatigue
- Properly balanced test material
- Randomization to prevent order effects

Subjective Listening Test Example 1

- Single stimulus
- Absolute category rating (ACR)
- Mean opinion score (MOS) scale

  Please rate the speech quality

  5=Excellent

  4=Good

  3=Fair

  2=Poor

  1=Bad

- Best suited for wide ranges of speech quality

Subjective Listening Test Example 2

- Dual stimulus (takes longer than single stimulus)
- Degradation category rating (DCR)
- Degradation mean opinion score (DMOS) scale

  Please rate the degradation of the second sample

  5=Imperceptible

  4=Perceptible but not annoying

  3=Slightly annoying

  2=Annoying,

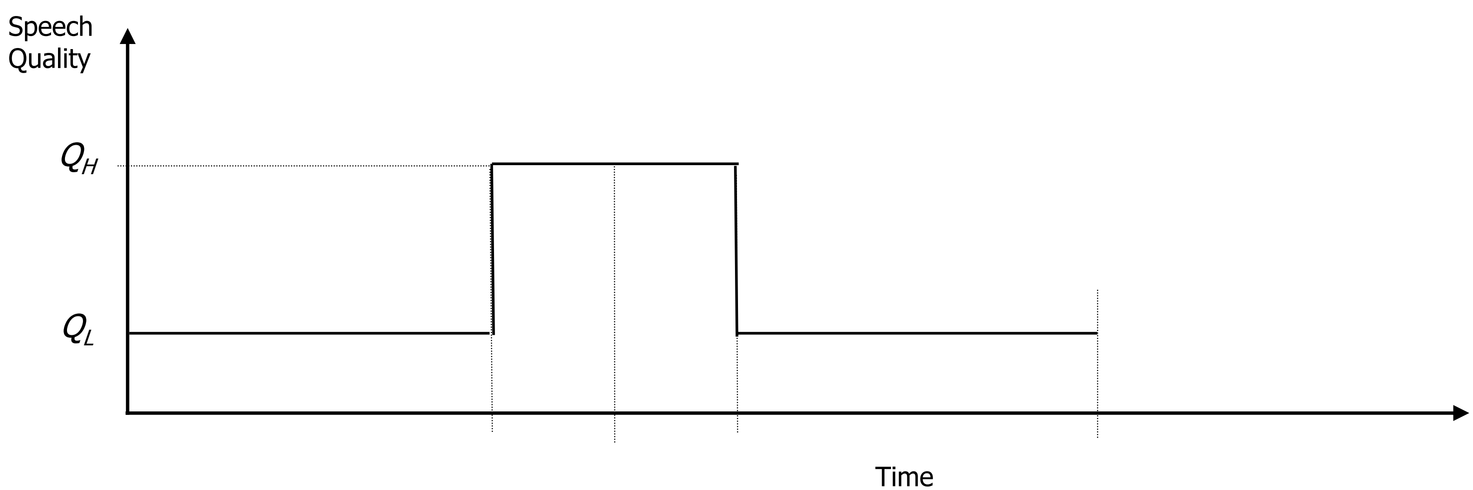
  1=Very annoying
- Scale mixes perception and opinion
- Able to resolve smaller quality differences

Issue - Time Varying Speech Quality

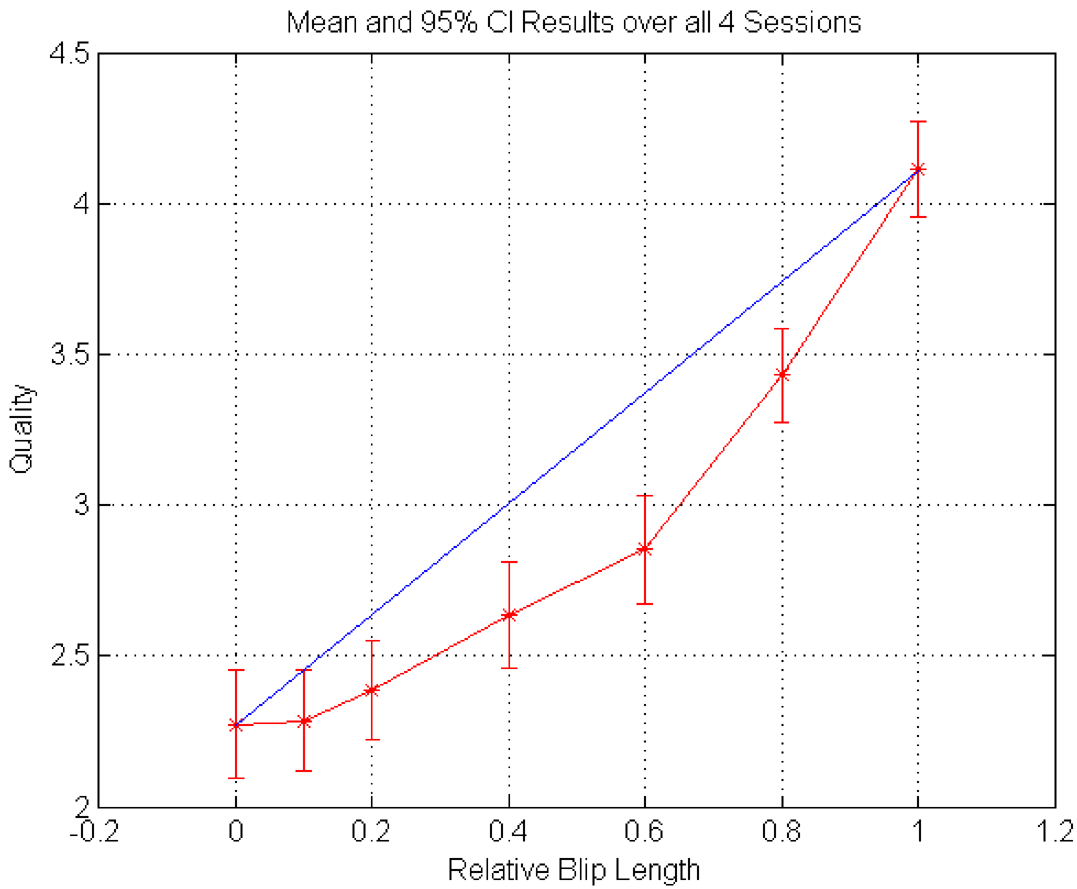
- Speech Quality in Telecom is no longer constant (e.g., wireless, VoIP)
- If quality varies moment by moment, what is the "overall quality"?

A Subjective Test

- A burst of higher quality in the middle of a ≈3 second recording
- Burst duration of 0 to 3 seconds
- Burst of lower quality also considered



A Key Result



$Q_L \approx 2.25$

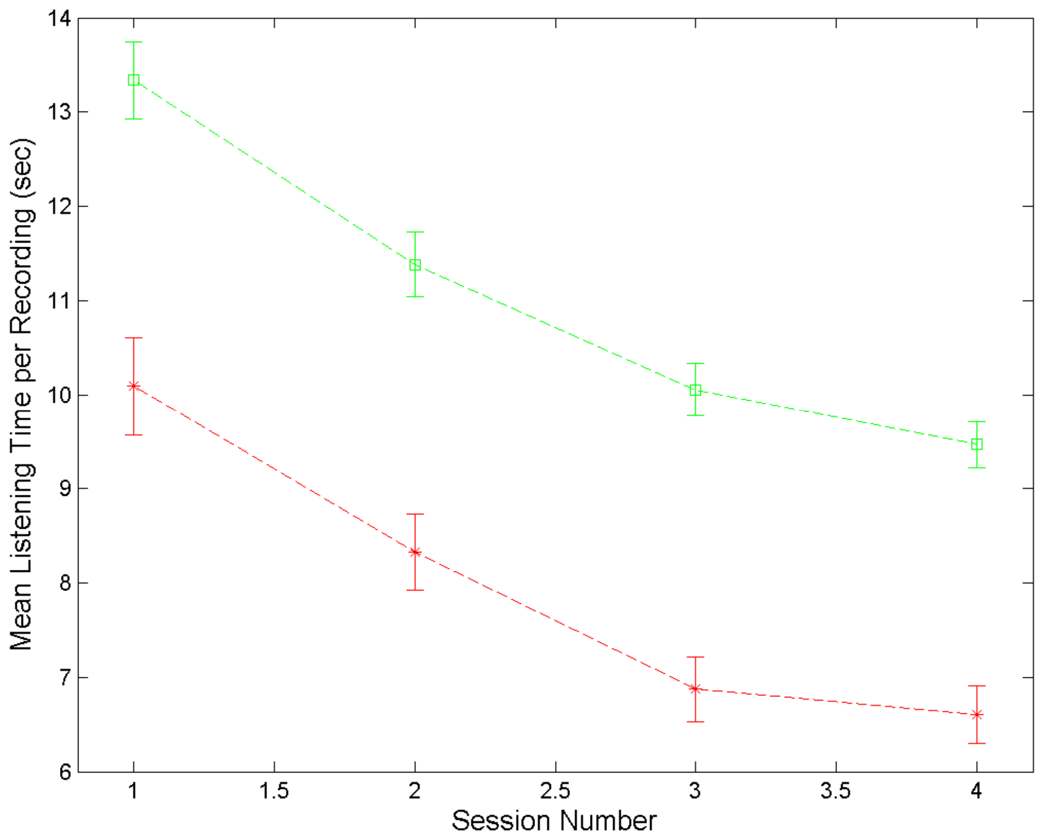
$Q_H \approx 4.10$

Mathematical mean (blue line) is too optimistic!

Issue – Fixed vs. Free Timing

- Most subjective tests force timing:
  - Hear 5-9 second recording
  - Rate its quality
  - Repeat
- Efficient implementation, control
- When do subjects naturally form opinions?
- ITS did pass/fail speech quality test, subjects could vote at *any* time after recording started

A Key Result



Green: listening times before "pass"

Red: Listening times before "fail"

Subjects are quick to criticize, slow to approve

Subjects speed up

Objective Estimation of Audio Quality

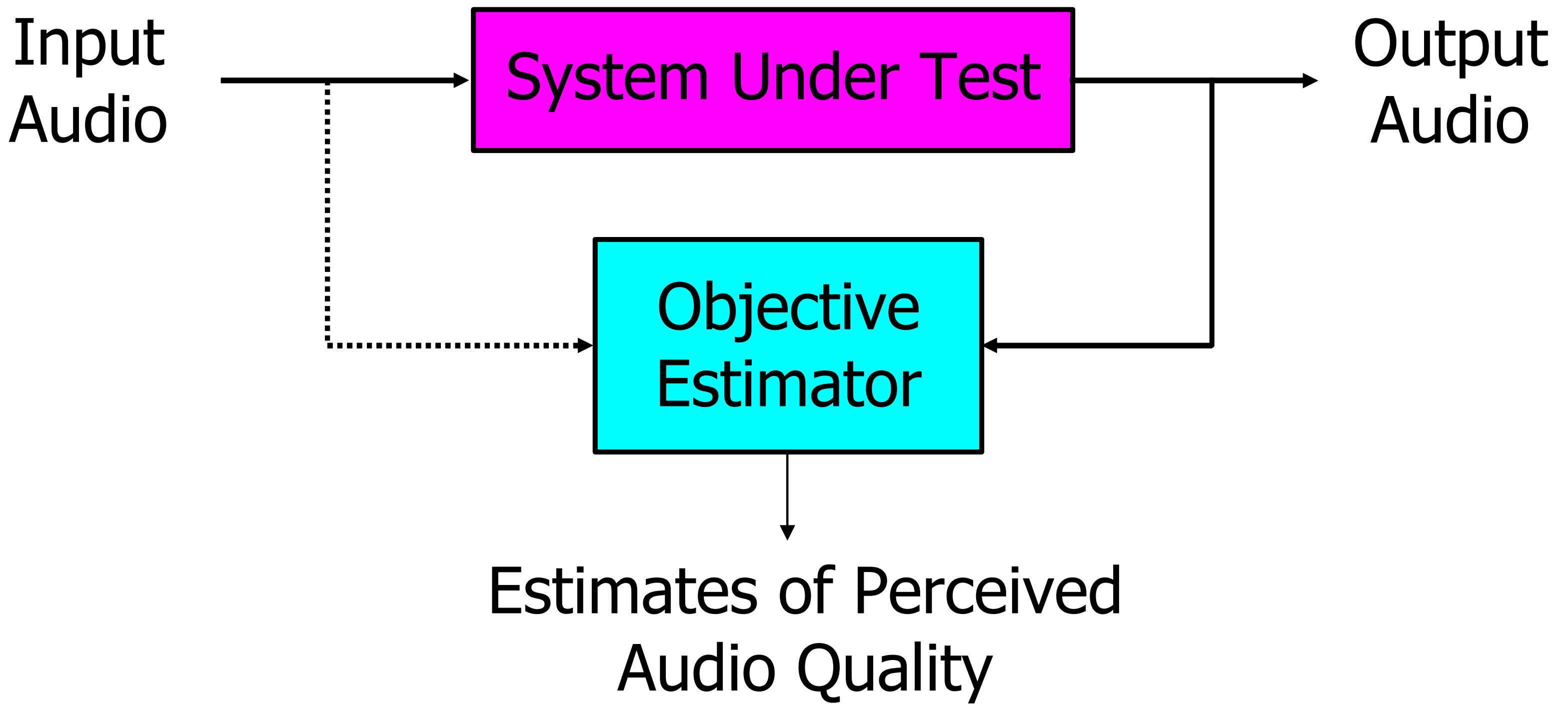
3 Main Approaches

- Telecom network element based (E-model)
- Measured network parameter based
- Audio signal based: DSP analysis of (sent) and received audio signals

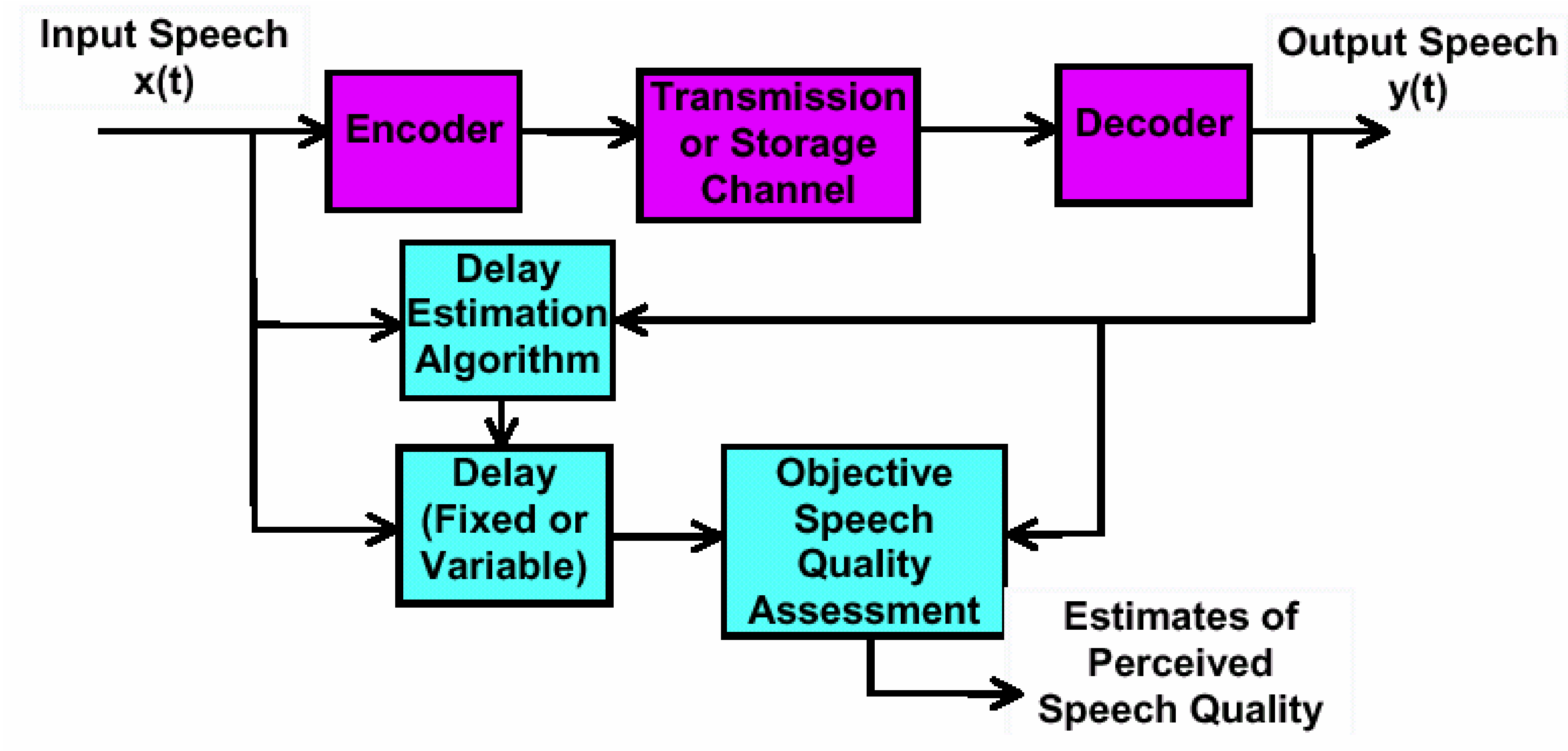
Objective Estimation vs. Subjective Measurement

- + Fast, inexpensive and repeatable
- + Just need software and some common hardware (e.g. off the shelf PC)
- + Listener variable and all associated issues are removed
- + Results are absolute, reference conditions not required
- - Can only *estimate* perceived audio quality

DSP Analysis of Audio Signals



Details on Speech or Audio Quality Estimation



Nomenclature

- Algorithms that require input and output are sometimes called
  - Input/output
  - Two-ended
  - Full Reference
  - Intrusive
- Algorithms that require just output are sometimes called
  - Output only
  - Output based
  - Single-ended
  - Zero Reference
  - Non-intrusive
- Trade-off
  - Simplicity (output only) vs. Accuracy (input/output)

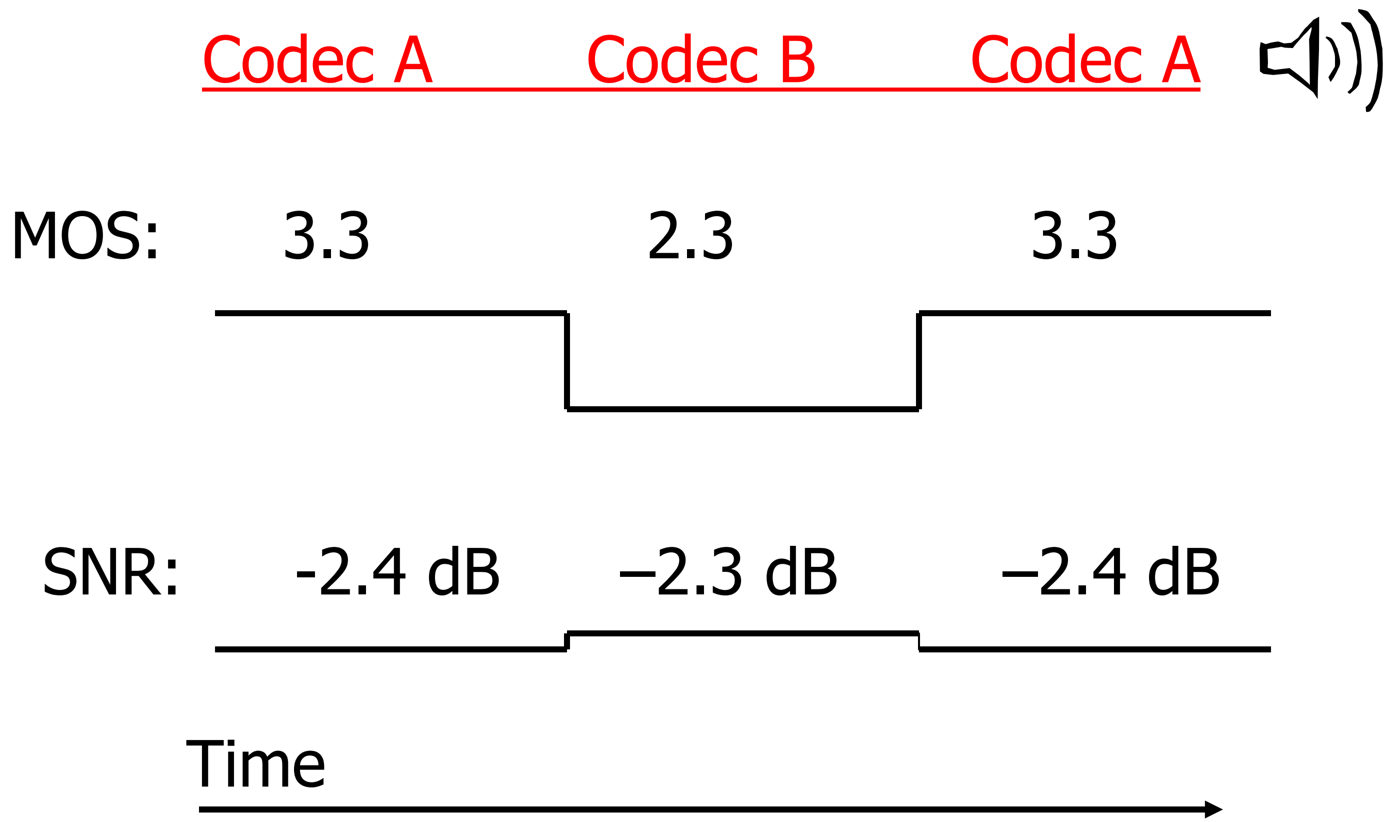
Use Waveforms to Estimate Quality?

- e.g., Segmental SNR

$$\mathrm{SNRseg}_j = \frac{1}{N}\sum_{i=0}^{N-1} 10 \cdot \log_{10}\left(\frac{x_{j\cdot N+i}{}^2}{(x_{j\cdot N+i} - y_{j\cdot N+i})^2}\right), \mathrm{SNRseg} = \frac{1}{M}\sum_{j=0}^{M-1} \mathrm{SNRseg}_j$$

  - Can measure coding noise or quantization distortion, each of which is related to audio quality
  - Does not measure perceived audio quality in general
  - Waveform fidelity is sufficient but not necessary for good audio quality

Example: Speech Coding

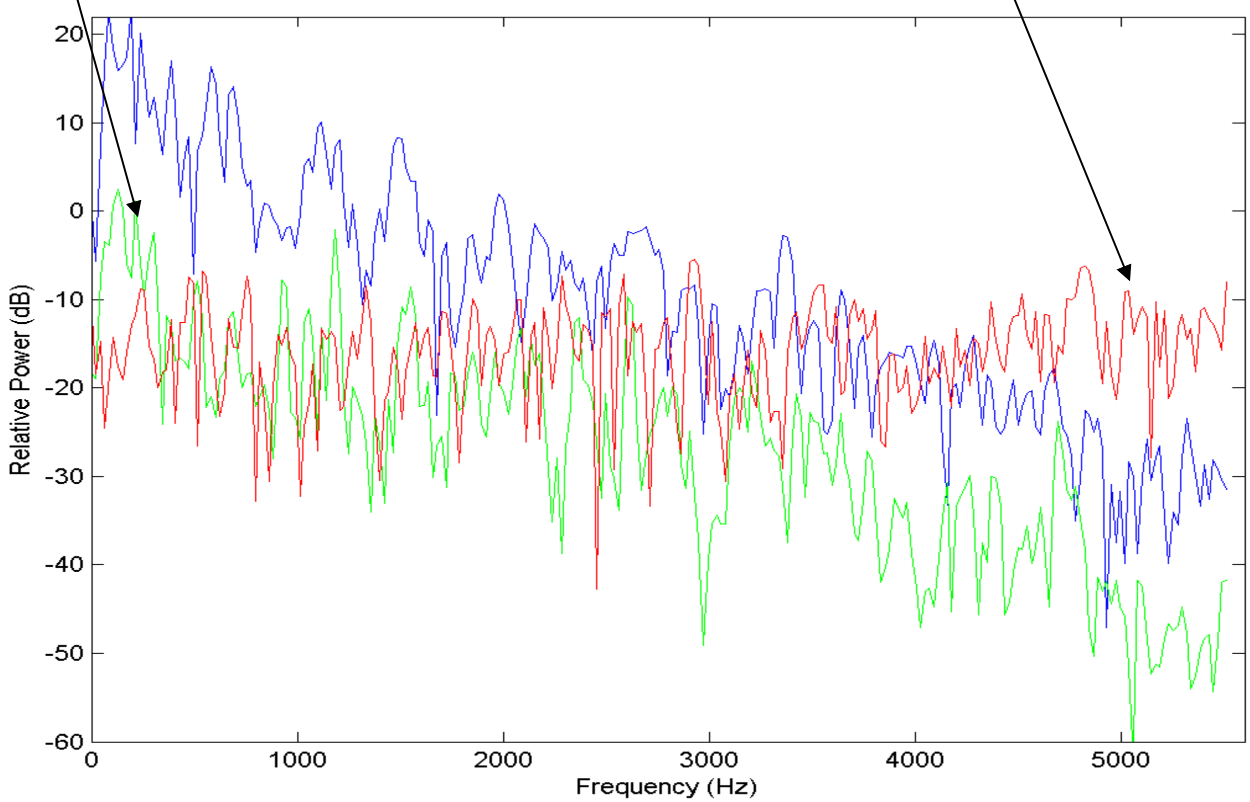


Parametric Speech Modeling

(not Waveform Coding)

Example: Music Coding

Codec 1: SNR=20 dB   Codec 2: SNR=20 dB



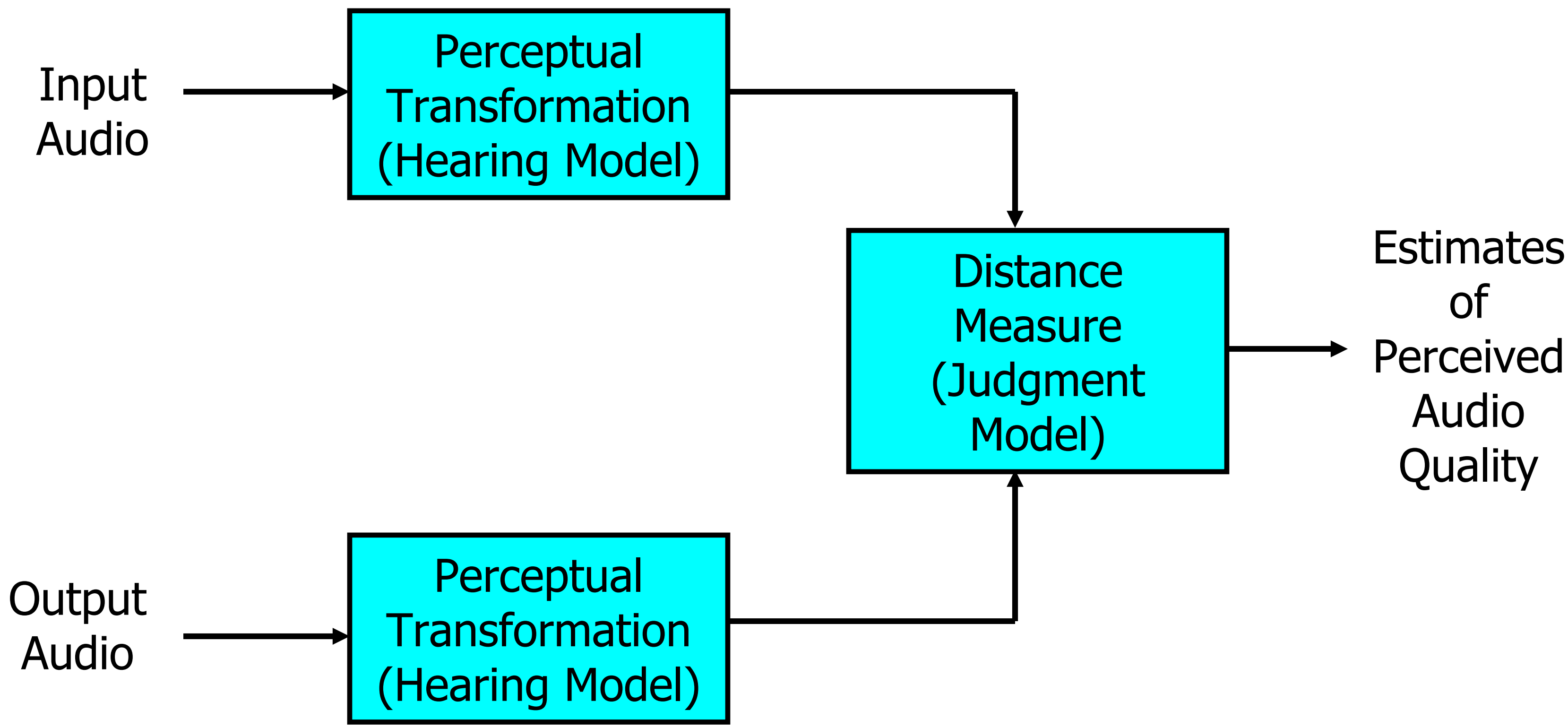
Perceptual Coding

(not Waveform Coding)

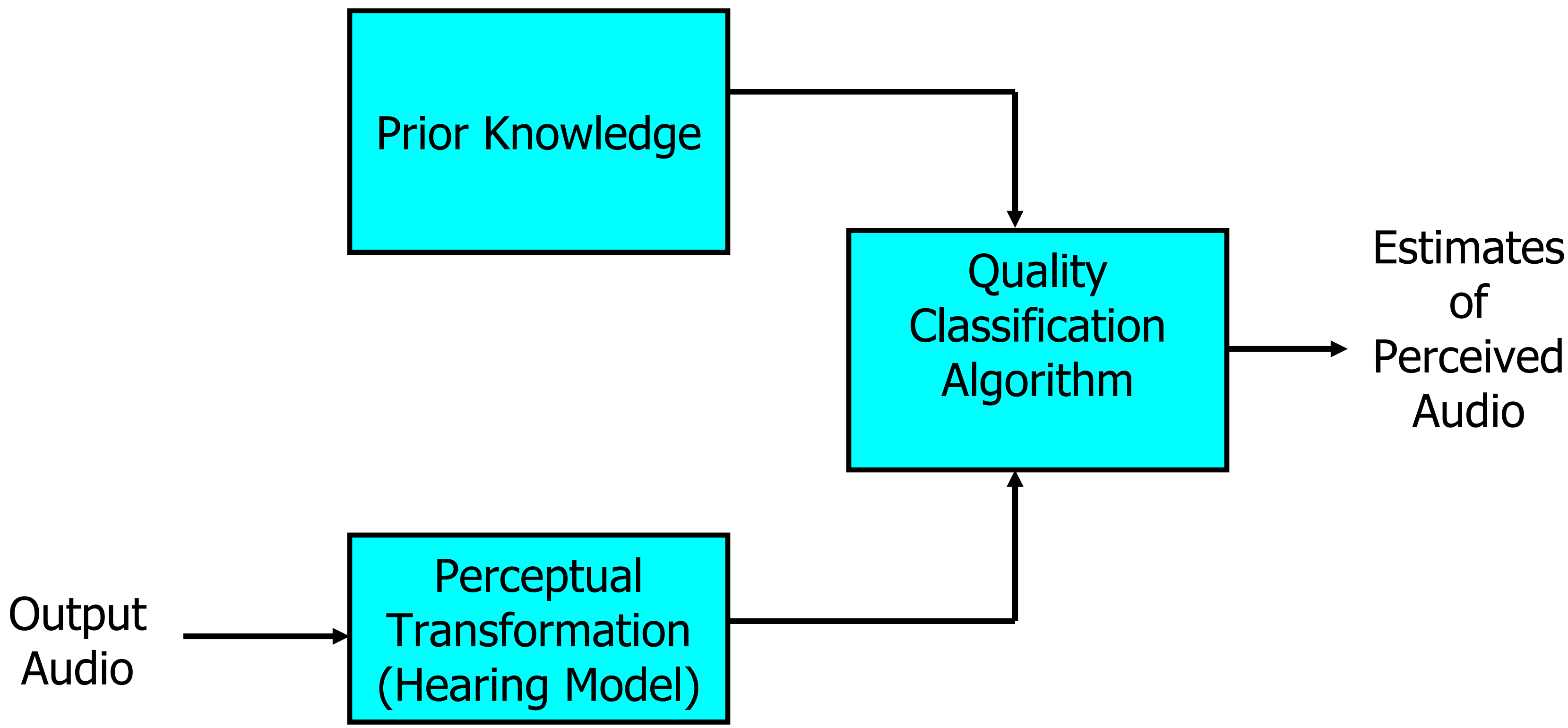
What to Do?

- Waveforms are all we have, yet waveforms do not tell the whole story
- Emulate what humans do
- Perception-based approach

Perception-Based Approach, Input/Output



Perception-Based Approach, Output-Only



Perceptual Transformation (Hearing Model)

- Frequency resolution not uniform on Hz scale $\Rightarrow$
  Use psychoacoustic frequency scale
  (Bark or Critical Band Scale)
- Loudness growth $\Rightarrow$
  Transform signal power to perceived loudness
- Sensitivity vs. frequency $\Rightarrow$
  Apply equal loudness weightings
- Hearing thresholds $\Rightarrow$
  Discard signals below threshold
- Frequency and time domain masking $\Rightarrow$
  Simulate by spreading signals in frequency and time

ITS Results (4 KHz Speech)

Based on experiments involving objective-subjective comparisons

- Distance measure (judgment model) is *at least* as important as perceptual transformation (hearing model)
- Some perception-based objective assessments may over-emphasize perceptual transformations (hearing model) and under-emphasize distance measure (judgment model)
- Hearing properties for noise and tone have been well-established through decades of research
- Judgment is relatively open question
- ITS contribution to problem involves more insightful and effective distance measures.
  - Measure and remove spectral deviations at one time or frequency scale then proceed to the next scale
  - Work from larger to smaller scales, because this is most likely to emulate listeners' patterns of adaptation and reaction

Standardized Algorithms

- Telephone Band Speech (300-3400 Hz), one talker, limited background noise
  - Input/Output
    - ITU-T P.862 (PESQ)
    - ANSI T1.518 (MNB)
  - Output only
    - ITU-T P.563
    - ANSI T1.??? (almost finished) (ANIQUE+)
- Wideband Speech (50-7000 Hz), one talker, limited background noise (input/output)
  - P.862.2 (Wideband PESQ)
- Full Bandwidth Music (20-20,000 Hz), Small Distortions (input/output)
  - ITU-R BS.1387 (PEAQ)

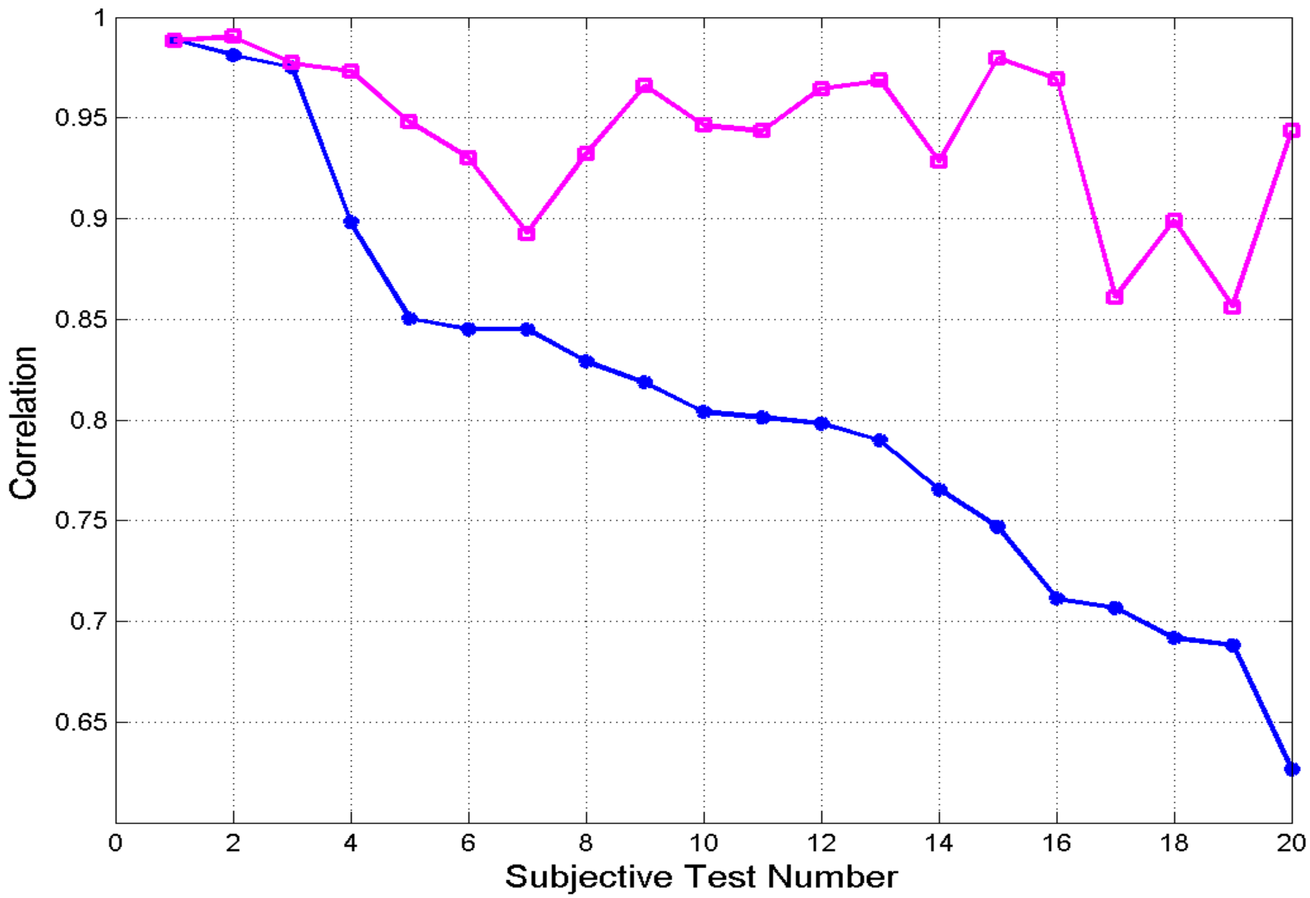
How Well Do They Work?

- Use subjective test results as “truth data”
- Look at objective-subjective correlation
- Look at objective-subjective RMS error

Example Results r.e. 20 Subjective Tests

Objective-Subjective Correlations (per condition)

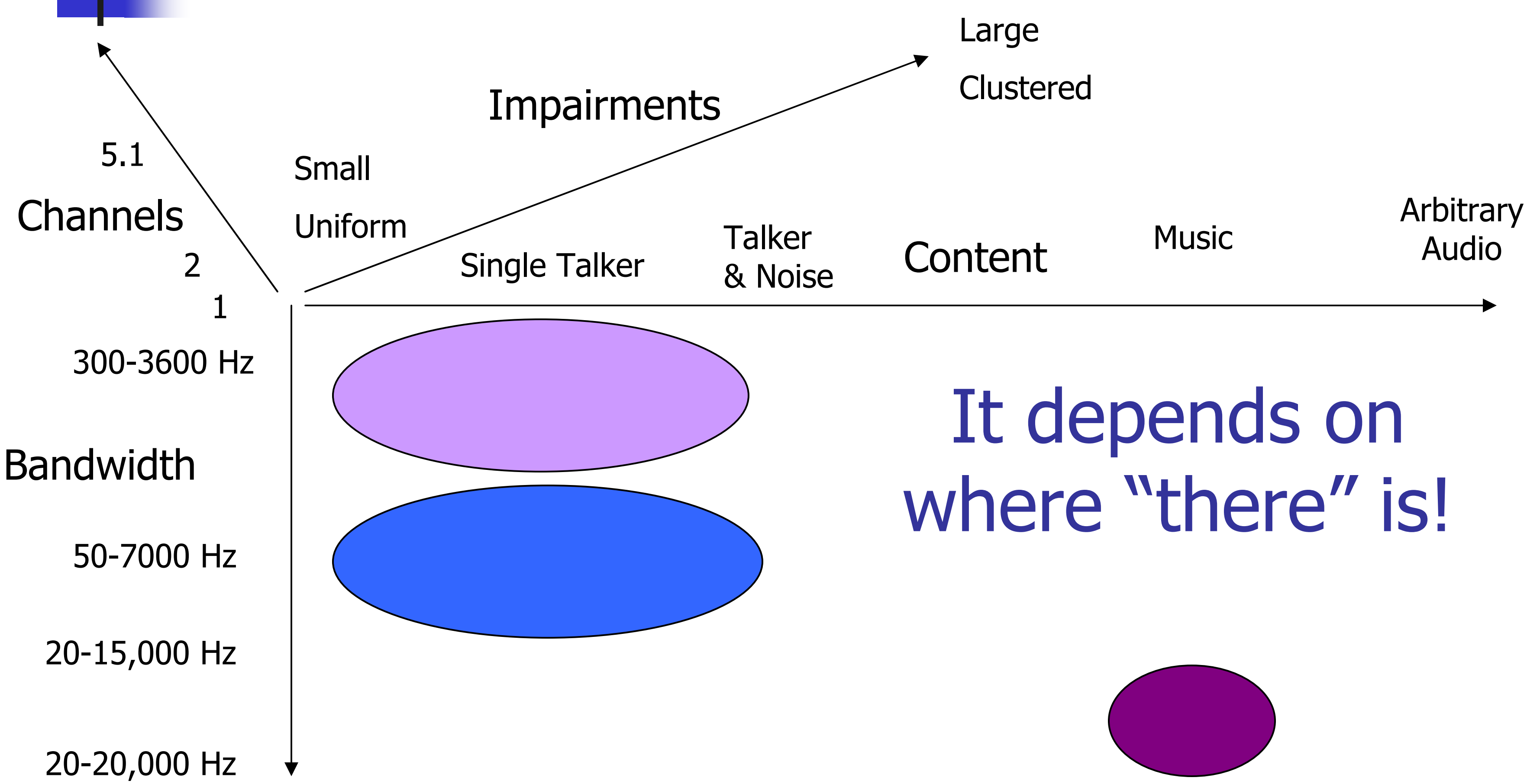
Telephone bandwidth (4 kHz) speech, single talker, limited background noise



P.862 (Input/Output)

P.563 (Output Only)

Are We There Yet?



It depends on where “there” is!

Moving Forward

- Reduced estimation error (esp. in output only algorithms)
- Increased applicability
  - Tandems of coders
  - Signal content (multiple speakers, speech + background noise, sound effects, music)
  - Signal bandwidth (7 kHz, 15 or 20 kHz)
- ITU-T SG12, Q9, P.OLQA
  - Wideband speech (50-7000 Hz)
  - Handsets & Handsfree
  - Music on hold
- ITU-R WP6Q, extend BS.1387
  - Add channels
  - Lower quality music coding

Objective Measurement of User-Perceived Video Quality

Stephen Wolf and Margaret Pinson
National Telecommunications and Information Administration (NTIA)
Institute for Telecommunication Sciences (ITS)
Boulder, CO

Outline of Presentation

- New Measurements Required
- Subjective Video Performance
- Objective Video Performance
- Video Standardization Efforts
- Extensions of Technology
- Demonstration of Tools

Estimating the User Quality of Service (QoS) Experience

- Two Automated Methods
  - Indirect Method - Measure Network Performance => Relate to User Experience
    - Difficult to Map to User Experience
    - Map is Dependent on Coder/Decoder
  - Direct Method - Measure Data Received by User => Relate to User Experience
- This Presentation – Direct Method

Advantages of Direct Method

- Measurement System "Sees" Exactly What User Sees
  - Important Because User QoS is User-Data Dependent
  - Scene Complexities (e.g., Spatial, Temporal) Significantly Influence Quality

Advantages of Direct Method

- Measurement System Can Be Made Technology Independent
  - Coder/Decoder Design
    - e.g., Error Concealment in Decoder
  - Transmission Method
- To Be Accurate (i.e., Track Subjective), In-Service Measurements are Required

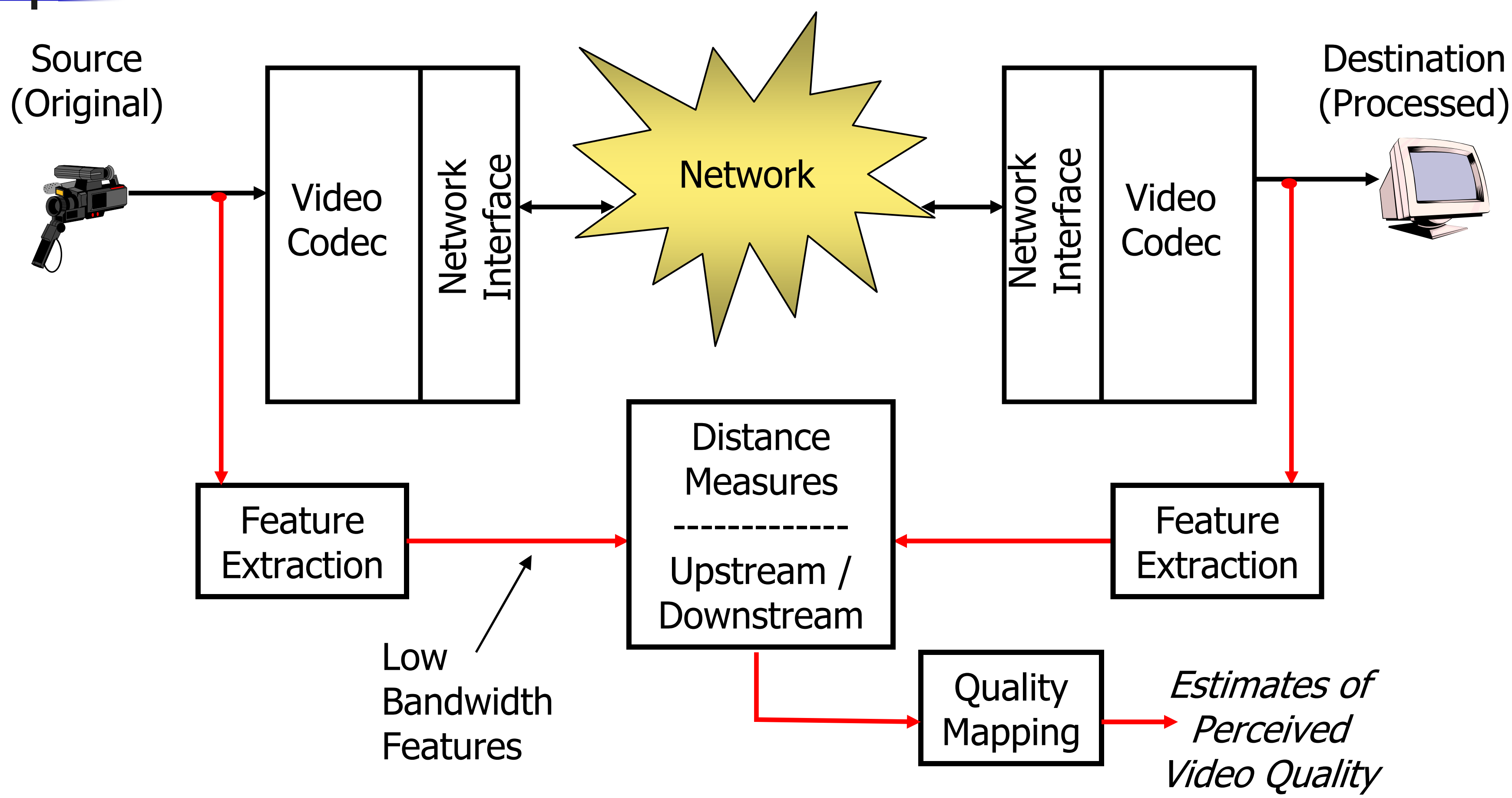
Three Direct Objective Methods (ITU-T Recommendation J.143)

- Full Reference (FR)
  - Full Access to Source Video
  - Accurate But Not Useful for In-Service
  - Scene and Technology Independent
- No Reference (NR)
  - No Access to Source Video
  - In-Service But Not Accurate
    - "Like Measuring Voltage without a Ground Wire"
  - Not Scene or Technology Independent

Three Direct Objective Methods (ITU-T Recommendation J.143)

- Reduced Reference (RR)
  - Uses Low Bandwidth Quality Features Extracted From Source and Destination Video Streams
  - In-Service Monitoring
  - Scene and Technology Independent
  - Accurate as FR, Degrades Gracefully as RR Bandwidth is Reduced

Reduced-Reference (RR) Measurement Paradigm



Outline of Presentation

- New Measurements Required
- Subjective Video Performance
- Objective Video Performance
- Video Standardization Efforts
- Extensions of Technology
- Demonstration of Tools

Subjective Testing Methods

- Truth Data for Objective Methods
- Standardized Test Procedures
  - ITU-R BT.500 (TV)
  - ITU-T P.910 (Multimedia)
- Controlled Test Environments
  - Monitor Setup, Viewing Distance, Lighting, etc.
- Controlled Test Methods
  - Training
  - Subject and Material Selection, Presentation, Scales

Subjective Testing Methods

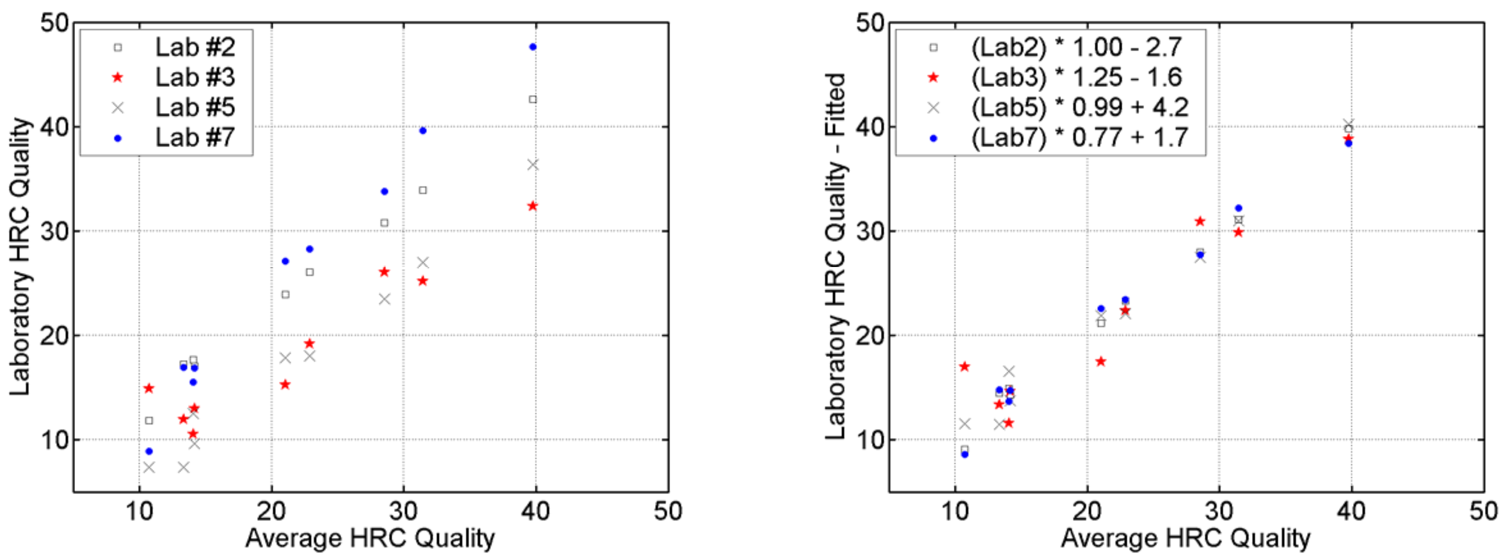
- Single Stimulus (SS) vs. Double Stimulus (DS)
  - Explicit or Hidden Reference for DS
  - DS More Robust (Context, Bias)
- Discrete vs. Continuous Assessment
- Randomization and Balance Over Test Variables is Very Important

Absolute Subjective Ratings

- For Practical Reasons, They Don't Exist!
  - Viewer Pool Variability, Criticality, Cultural Differences
  - Overall Range of Quality in Experiment
  - Application Dependencies
  - Laboratory Setup and Methodology
- Relative Ratings are Stable

Absolute Subjective Ratings

- Video Quality Expert's Group (VQEG)
  - Phase 1, *Identical* Experiment, 4 Labs
  - Significant Gain and Offset Issues



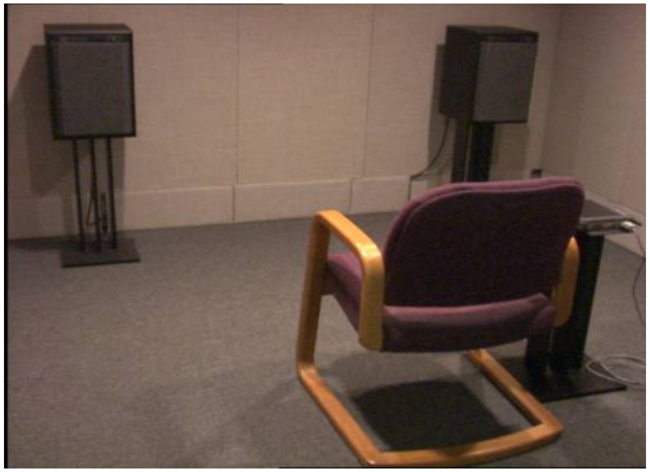
ITS-Developed Data Mapping

- Observations Led to the Development of Mathematical Data Mapping Method
- Iterative Nested Least Squares Algorithm (INLSA)
  - Linear Transformation of Subjective Data Sets
  - Simultaneous Minimization of Objective Estimation Error
- Produces Large Coherent Subjective Data Base
  - Better Video Quality Models (VQMs)

Recent Papers on INLSA and Subjective Video Testing Issues

- July, 2003, "An Objective Method for Combining Multiple Subjective Data Sets," SPIE Video Communications and Image Processing Conference, Lugano, Switzerland.
- July, 2003, "Comparing Subjective Video Quality Testing Methodologies," SPIE Video Communications and Image Processing Conference, Lugano, Switzerland.
- “spie03obj.pdf” and “spie03subj.pdf”
  - http://www.its.bldrdoc.gov/n3/video/documents.htm

ITS Subjective Test Facilities



- Two Sound Isolated Rooms
  - Interactive Tests
  - Configurable (A, V, A/V)
- Fully Automated
  - A/V Playback
  - Subjective Scores

ITS Subjective Test Facilities

Uncompressed HDTV Equipment

HDTV Viewing Room

ITS's Extensive Subjective Data Sets

- 19 SD Data Sets (2651 Video Clips) + 1 HD Data Set
- 392 SD Scenes
  - Wide Range of Spatial Detail, Motion, Contrast, Brightness, Color
- 272 SD HRCs (Video Systems)
  - Wide Range of Bit Rates, Coders, Transmission Channels, including Analog
- National Archive

Outline of Presentation

- New Measurements Required
- Subjective Video Performance
- Objective Video Performance
- Video Standardization Efforts
- Extensions of Technology
- Demonstration of Tools

Objective Video Quality-Estimation Components

- Step 0: Calibration
  - Remove Perceptually-Irrelevant Systematic Errors from Destination Signal
  - Report Separately From Perception-Based Quality
- Step 1: Feature Extraction
  - Extract Perceptually-Relevant, Quality-Related Attributes from Source and Destination Signals
  - .00001 < (Feature Bit-Rate/Signal Bit-Rate) < 1
    - "Reduced Reference" In-Service Measurements
- Step 2: Distance Measures
  - Calculate a Perceptually-Relevant Distance Between the Source and Destination Feature Streams
- Step 3: Quality Mapping
  - Relate Perceptual Distances to Estimates of Perceived Quality

Objective Video Quality-Estimation Components

- Most Step 0 to 3 Algorithms
  - Documented in NTIA Report 02-392, "Video Quality Measurement Techniques"
    - "vqm_techniques_v2.pdf" available at: http://www.its.bldrdoc.gov/n3/video/documents.htm
  - Sept, 2004, "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Transactions on Broadcasting,* v. 50, n. 3, pp. 312-322.
    - "ieee04.pdf" also available at: http://www.its.bldrdoc.gov/n3/video/documents.htm
- Additional Documentation Will Be Presented When Covered

Step 0: Calibration

- Estimate and Remove (Order is Important)
  - Temporal Shifts (Video Delay)
  - Spatial Scaling (Horizontal and Vertical)
  - Spatial Shifts (Horizontal and Vertical)
  - Valid Region Reduction
  - Gain and Level Offset
- Importance Depends Upon Objective Measurement Technique
  - PSNR is Highly Sensitive
  - RR Measurements are Much Less Sensitive

Step 0: Calibration

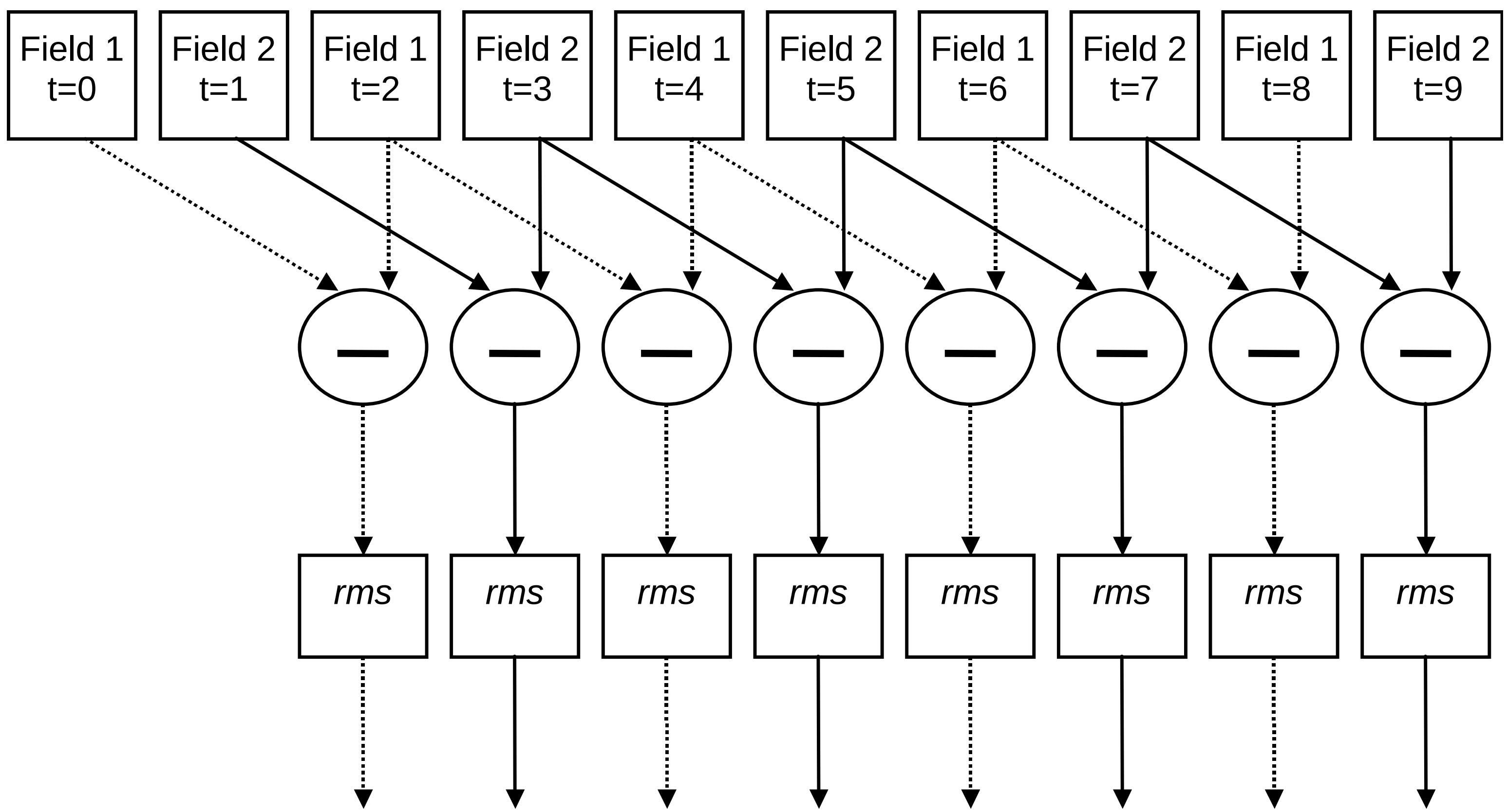
- Most Calibration Quantities are Fixed for a Given Hypothetical Reference Circuit (HRC) – i.e., system under test.
  - Except Temporal Registration
- Need Only Estimate Once or Infrequently at System Setup
- Spatial Scaling and Spatial Registration are CPU and Bandwidth Intensive

Temporal Registration – Two Methods

- Frame Based
  - Slow, High RR Bandwidth, Sensitive to Calibration
  - Laboratory, Bench-Top, Out-of-Service
  - Produces Delay Estimate for Each Frame
- Sequence Based
  - Fast, Low RR Bandwidth, In-Sensitive to Calibration
  - In-Service
  - Produces Delay Estimate for Each Sequence
  - Works Well with RR Quality Measurements

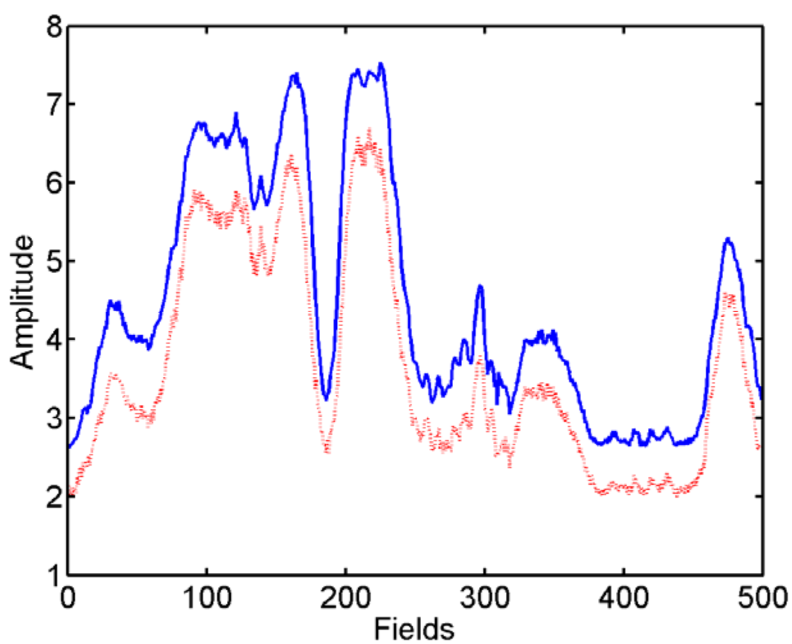
Sequence Based Temporal Registration

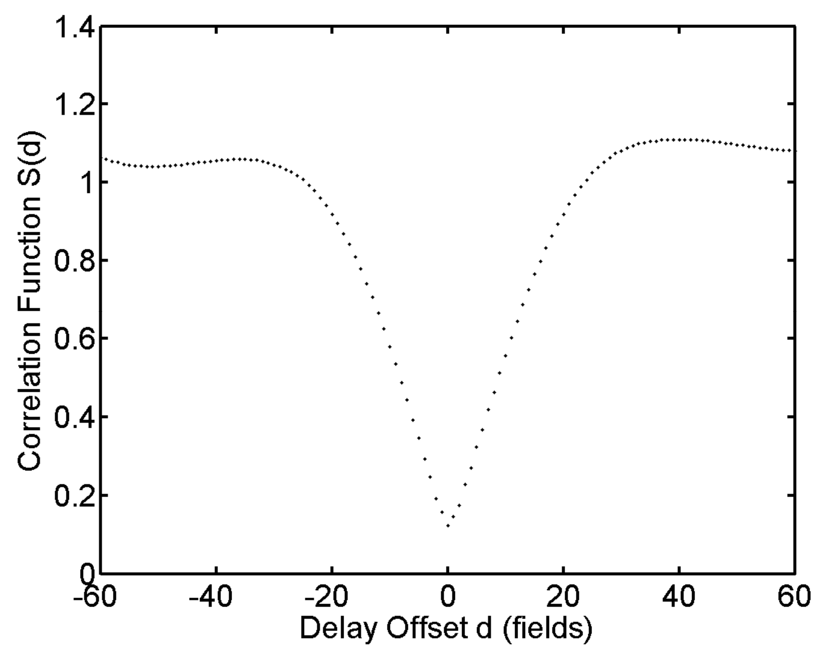
- Uses Motion, Luma (Y') Features



Sequence Based Temporal Registration

- Correlation to Align Source (Blue) and Destination (Red) Feature Streams





Spatial Scaling Estimation

- More Applicable to Multimedia than TV
  - Has Been Observed For TV (Rare)
- Uses H and V Image Profiles
  - Average of Image Columns (H), Rows (V)
  - Reduces Scaling Estimation Complexity
- Use Smart Search to Find Optimal Stretch or Shrinkage
- Use Several Scenes & Filter Results

Spatial Registration Estimation

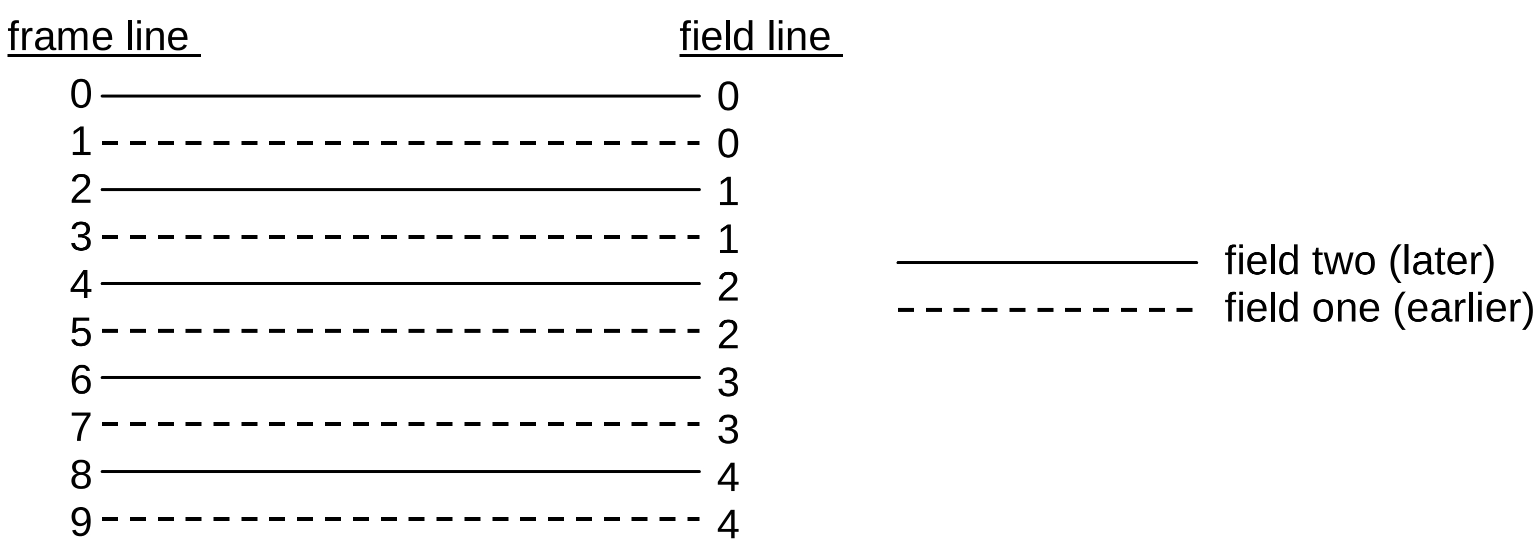
- Special Test Signals Can Be Useful for Out-of-Service Measurements

Spatial Registration Estimation

- In-Service Methods Must Utilize Real Video Scenes
- Multi-Dimensional Search (Expensive)
  - Utilize Temporal Registration Results
  - Random Sampling of Pixels (e.g., 0.2%)
  - Sub-Sample Temporally (e.g., every ½ sec)
- Use Several Scenes & Filter Results

Spatial Registration Estimation

- Spatial Registration Uses Field Processing
  - Some Video Systems “Re-Frame”



Spatial Scaling and Registration Estimation

- NTIA Technical Report TR-06-433, "Reduced Reference Video Calibration Algorithms"
- "ntia_tr_06_433.pdf" available at:
  - http://www.its.bldrdoc.gov/n3/video/documents.htm

Valid Region Reduction

- Video Systems Often Reduce Picture Area
  - Mechanism to Save Bits
- Use Mean Value of Columns & Rows
  - Transitions (Ramps) Not Included
- Referenced to Source, Accounting for Spatial Scaling and Shifts

Gain and Level Offset – Independent Color Components

- Temporally & Spatially Scaled & Registered
- Treat Each Component (e.g., Y', Cb, Cr) Separately
- Divide Valid Region into N Sub-Regions
  - Take Mean of Each Sub-Region
  - Solve Least Squares Problem

$$\underline{D} = g\underline{S} + l$$

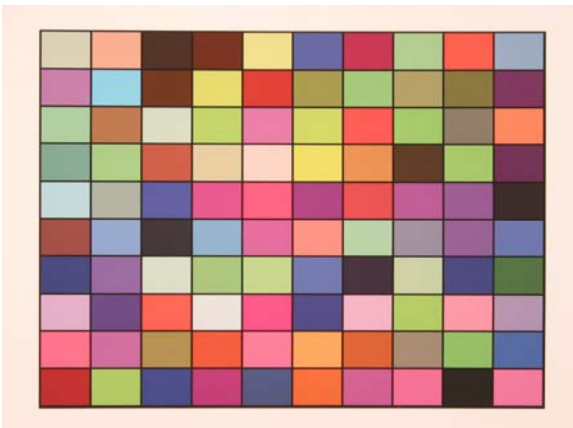
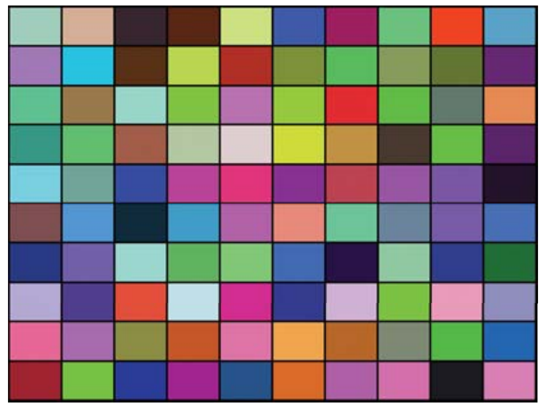
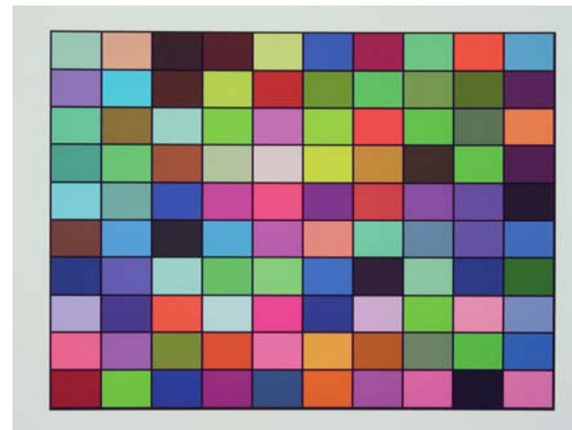
Gain and Level Offset – Dependent Color Components

- Required for Complex Color Errors (e.g., Hue)
- Estimate and Remove Arbitrary Color Component Mixing Plus DC Shift
  - Must Solve for 12 Unknowns
  - e.g., For RGB Color Space, Red Color Component of 1st Corrected Sample is

$$\hat{S}\_R_1 = A_{1,1} + A_{2,1} * D\_R_1 + A_{3,1} * D\_G_1 + A_{4,1} * D\_B_1$$

Gain and Level Offset – Dependent Color Components

Impaired | Original | Corrected



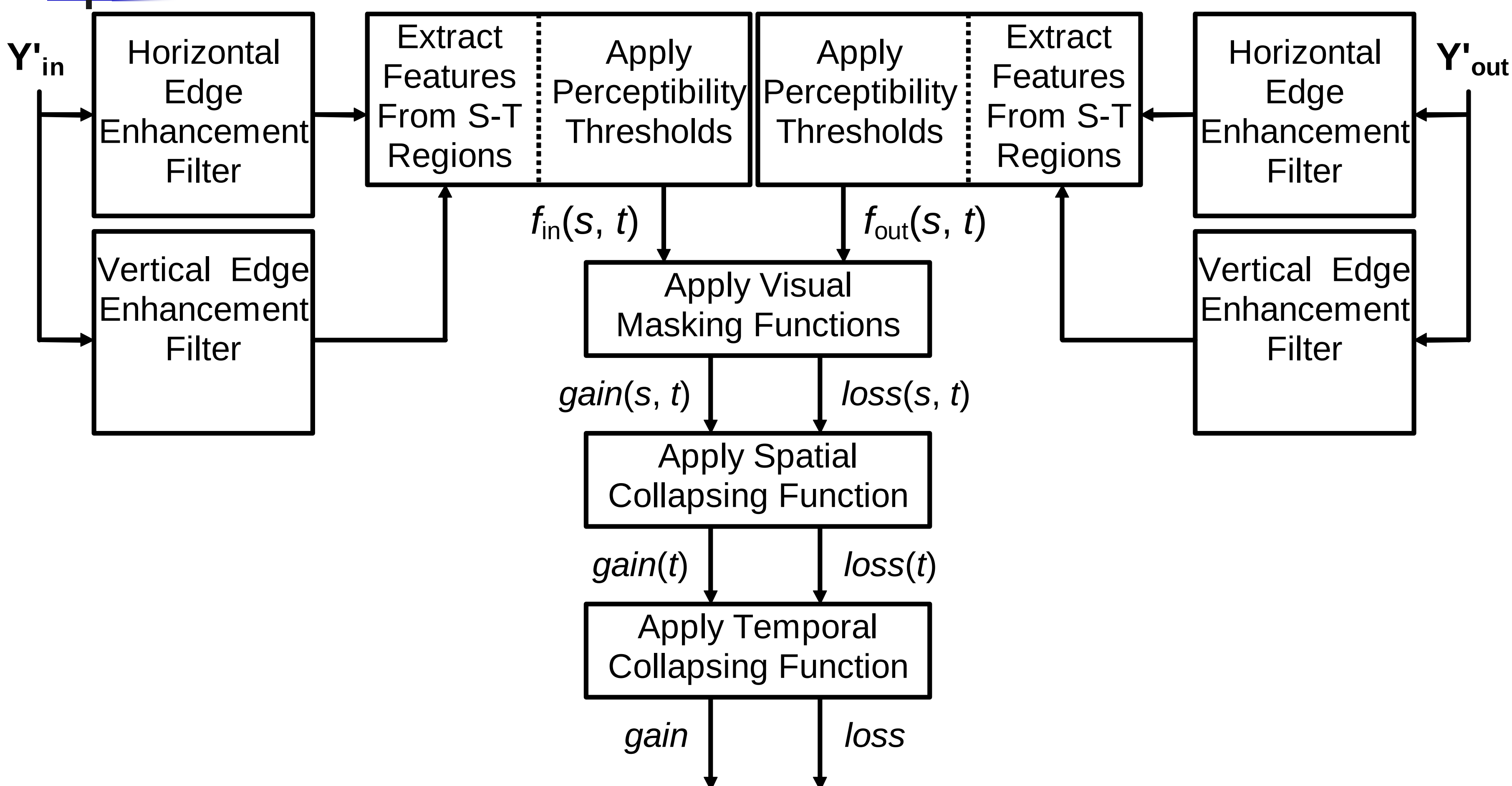
Gain and Level Offset – Dependent Color Components

- NTIA TM-04-406, “Color Correction Matrix for Digital Still and Video Imaging Systems”
- “ntia406.pdf” available at:
  - http://www.its.bldrdoc.gov/n3/video/documents.htm
- Excellent Practical Video Color Space Reference “ColorFAQ.pdf” available at:
  - http://www.poynton.com/PDFs/ColorFAQ.pdf

Step 1: Feature Extraction

- ITU-R Rec. BT.601 Sampling
  - 4:2:2, 13.5 MHz
  - Color-Difference (Cb, Cr) Half Bandwidth
  - Gamma Pre-Corrected (Approximates HVS)
- Luma (Y') Features Characterize
  - Spatial Gradients (Edges, Angles)
- Color-Difference (Cb, Cr) Feature
  - Treated as Two-Dimensional Vector
- Temporal Gradients (Motion) of Y', Cb, Cr

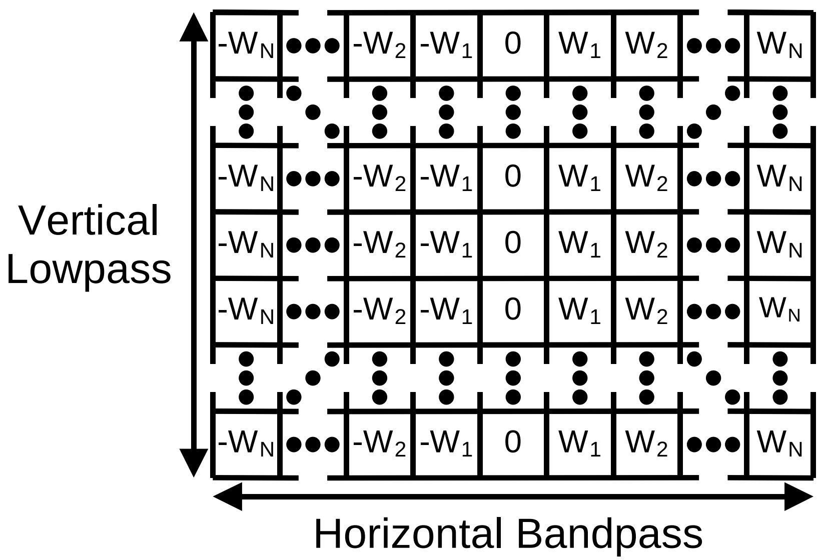
Overview of Spatial Gradient Feature Extraction

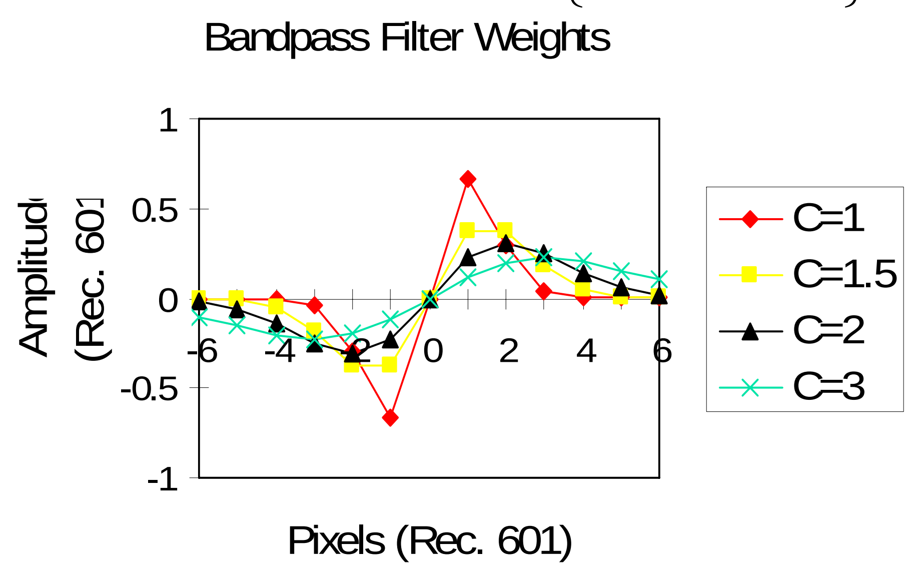


Spatial Gradient Filters

- Edge Enhancement Filters (C = 2)
- Lowpass / Bandpass

$$w_x = k * \left(\frac{x}{c}\right) * \exp\left\{ -\left(\frac{1}{2}\right)\left(\frac{x}{c}\right)^2 \right\}$$

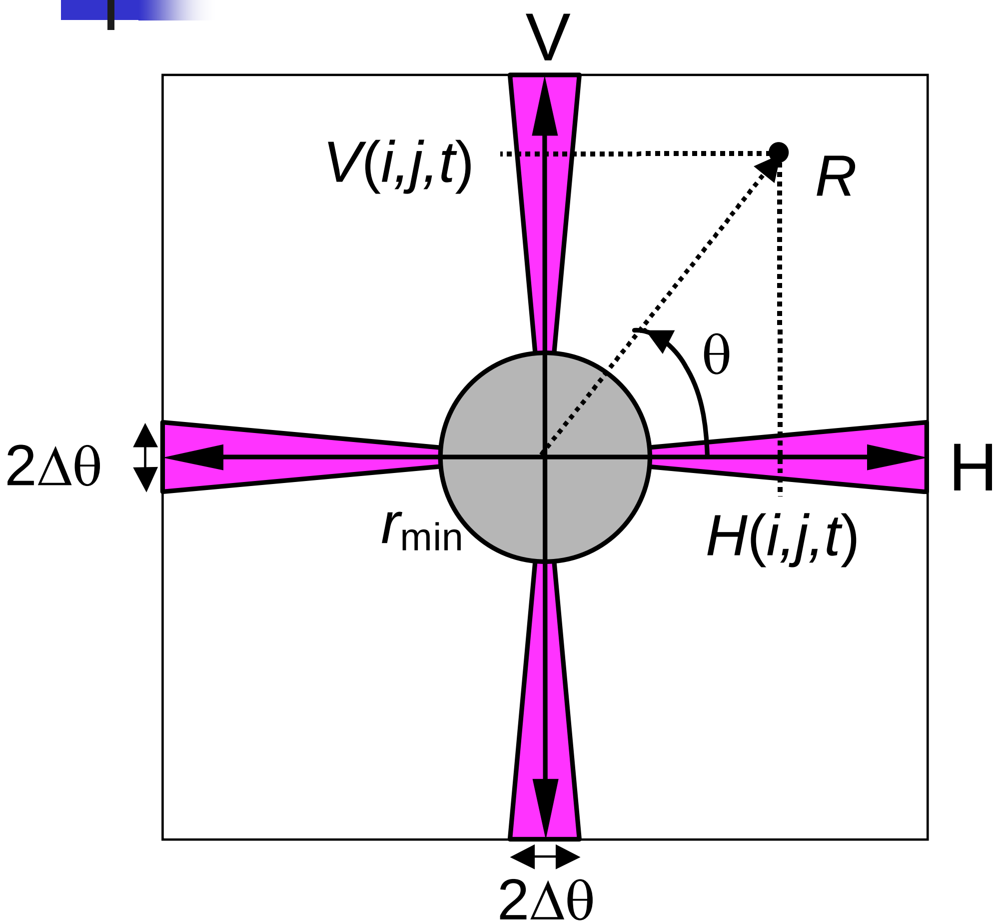




Two Spatial Gradient Features

- SI13
  - Spatial Gradient Magnitude
  - Blurring
- HV13
  - Spatial Gradient Angular Distribution
  - Blocking/Tiling

Graphical Depiction of Spatial Gradient Features



- SI13

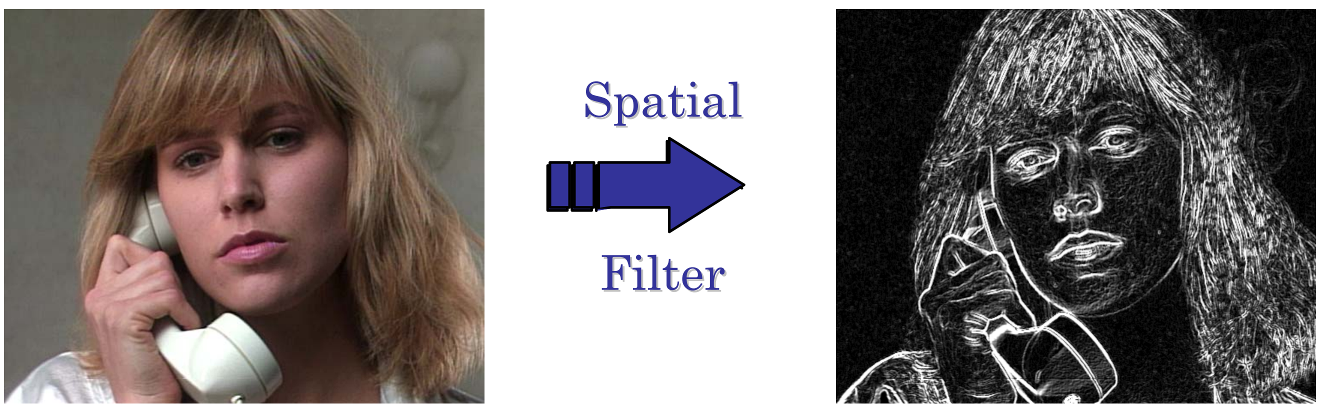
  $$\{\text{stdev}[R(i,j,t)]\}|_P$$

- HV13

  $$\frac{\{\text{mean}[R_{HV}(i,j,t)]\}|_P}{\{\text{mean}[R_{\overline{HV}}(i,j,t)]\}|_P}$$

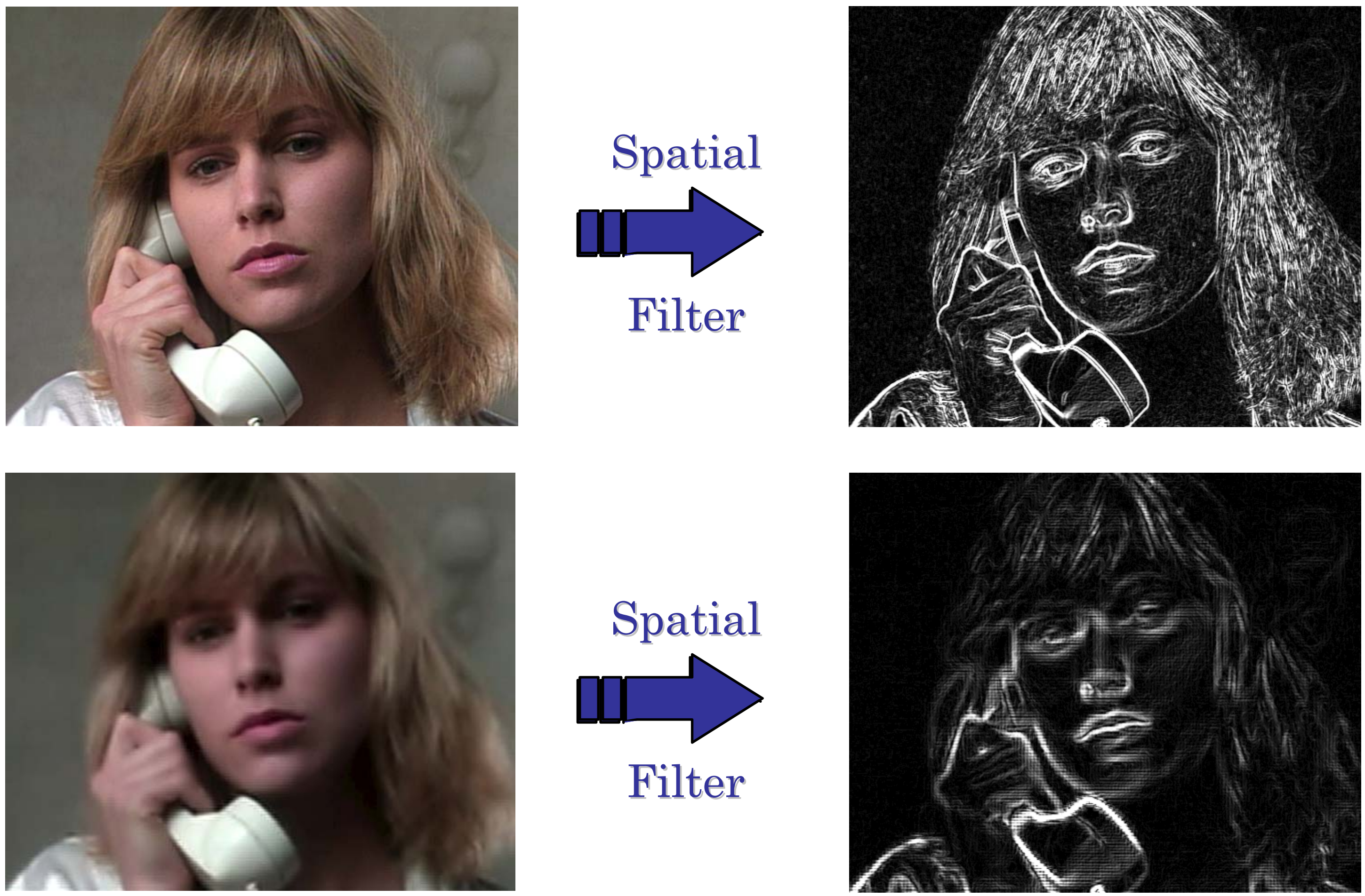
Spatial Gradient Feature Examples

- Missing Edges - Blurring, Smearing
- Added Edges - Tiling, Edge Busyness, Lines



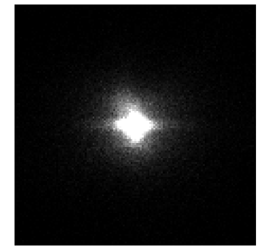
Spatial Information (SI)

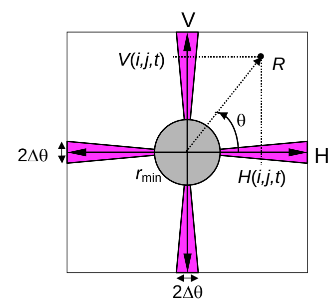
Blurring



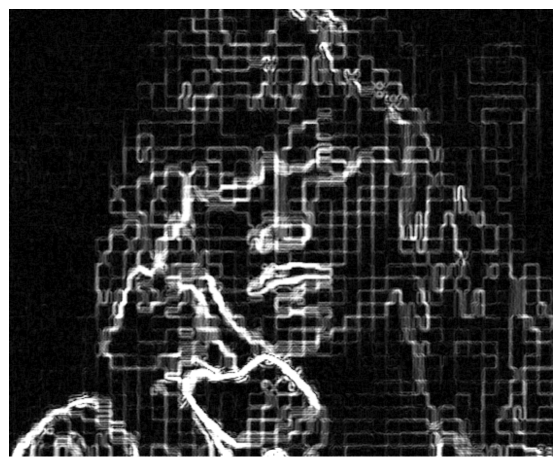
Tiling / Blocking

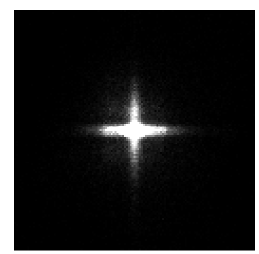
Spatial Filter





Spatial Filter





Temporal Gradient Features

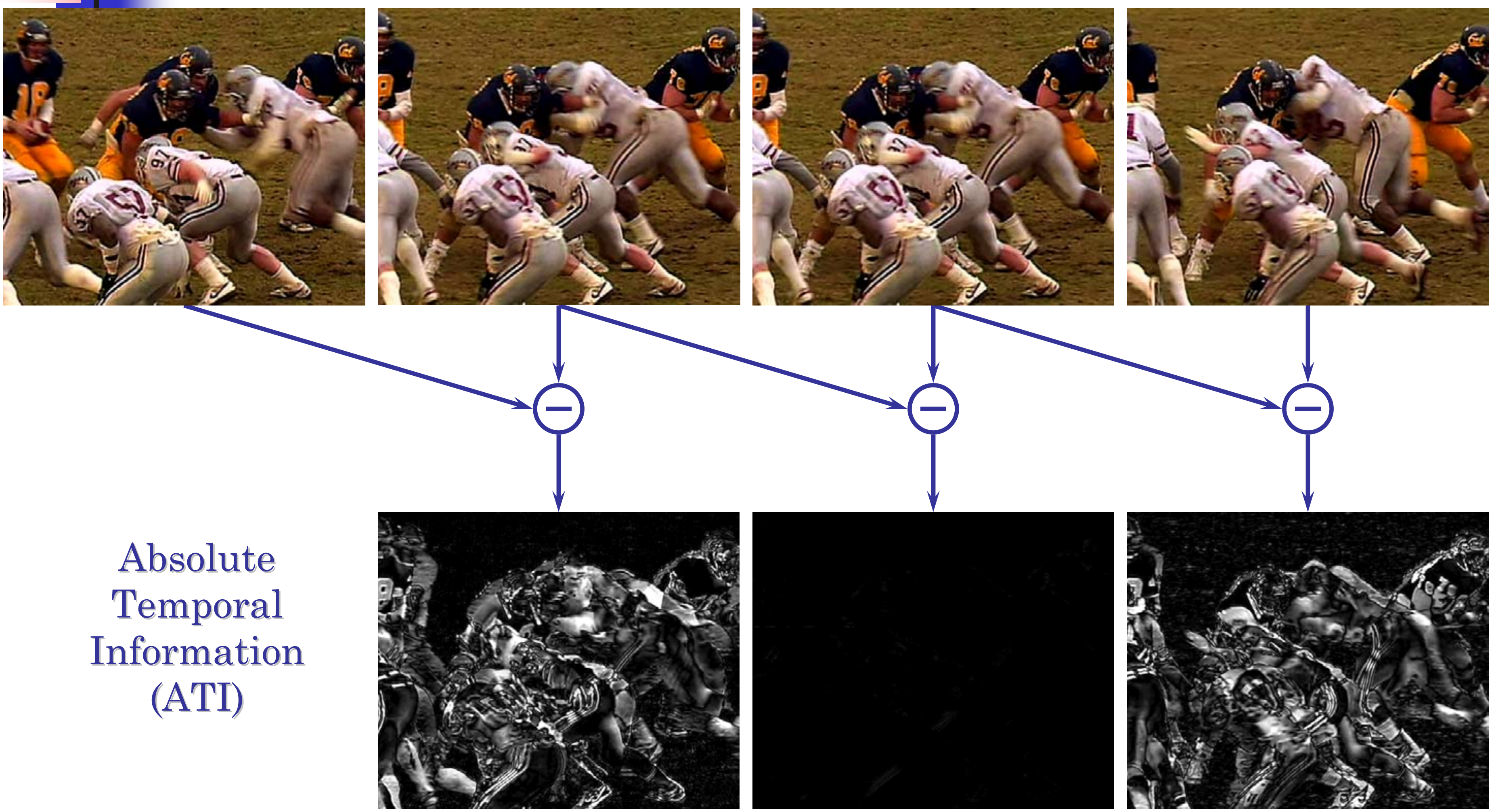
- Absolute Value of Frame Differences
- Missing Motion - Frame Freezes, Dropped Frames
- Added Motion - Error Blocks, Noise

Frame n

Frame n-1

Absolute Temporal Information (ATI)

Frame Freezes, Dropped Frames



Error Blocks

Frame n

−

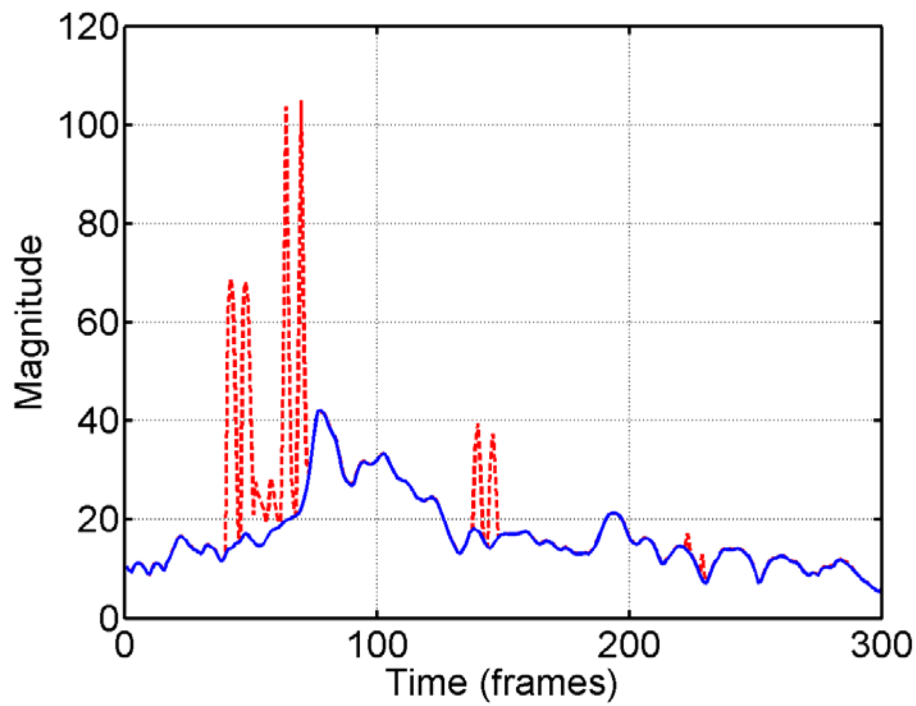
Frame n-1

=

Absolute Temporal Information (ATI)

Temporal Feature
Lowbw (10 kbits/s) VQM

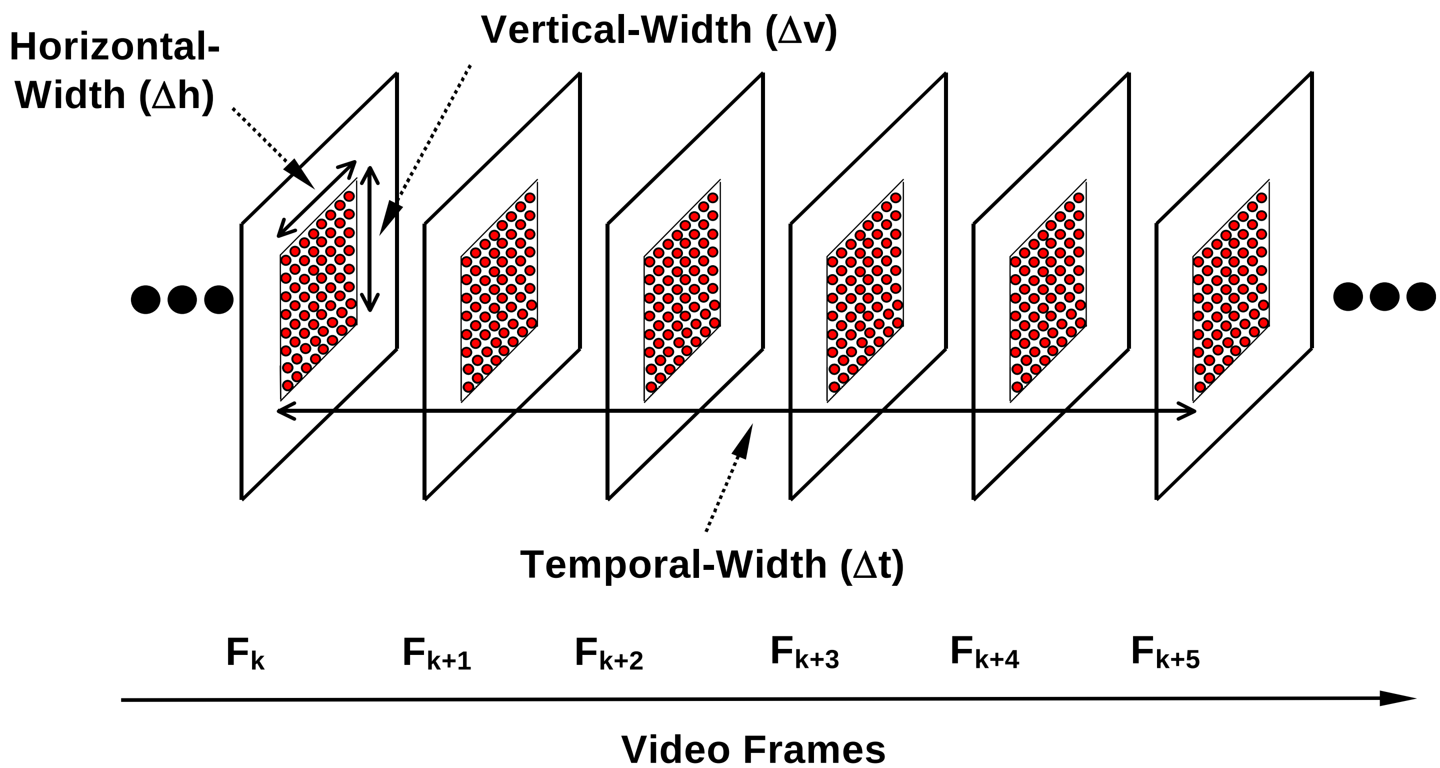
- For Noise & Errors
  - ATI = RMS { $Y'C_bC_r(t) - Y'C_bC_r(t\text{-}0.2s)$ }



0.2s Makes Feature Insensitive to Frame Repeats Present in Low Frame Rate Multimedia Systems

Feature Extraction From S-T Regions

- Summary Statistics Extracted from S-T Regions After Perceptual Filtering (e.g., Mean, Stdev)

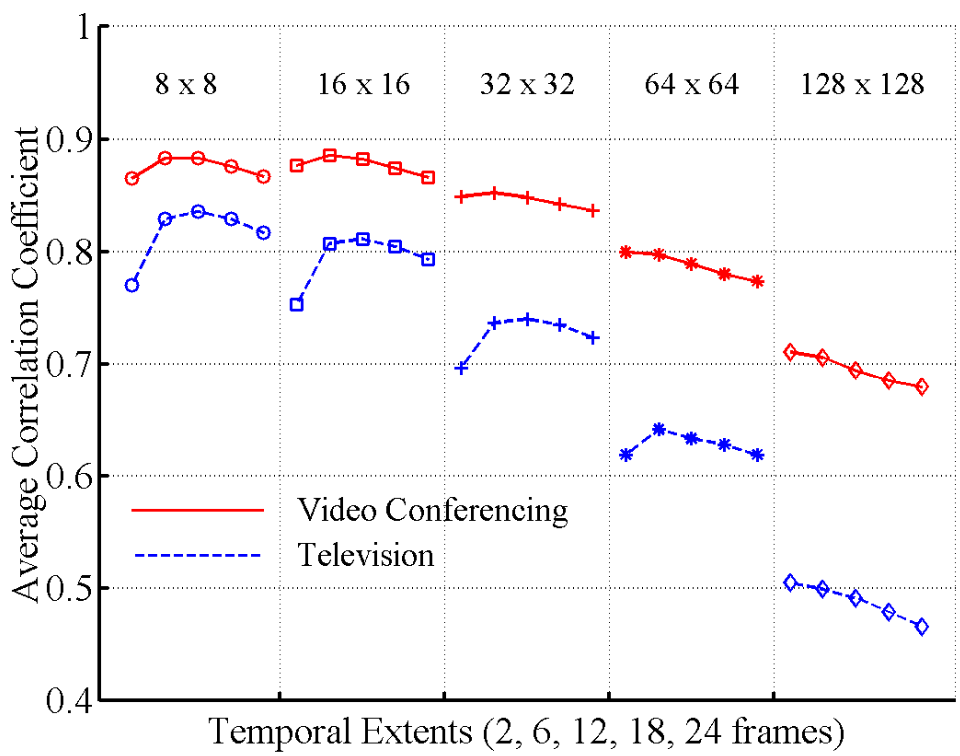


Objective to Subjective Correlation: S-T Region Size

- Optimal S-T Region Size
  - 8 x 8 x 0.2 Seconds (Used in General VQM)
  - Shorter Temporal Extents => Poorer Correlation
- Correlation Falls Off Slowly For Increasing Temporal Extents
  - 32 x 32 x 1 Second - Lowbw (10kbits/s) VQM

Objective to Subjective Correlation: S-T Region Size

- HV Loss (General VQM Error Pooling)



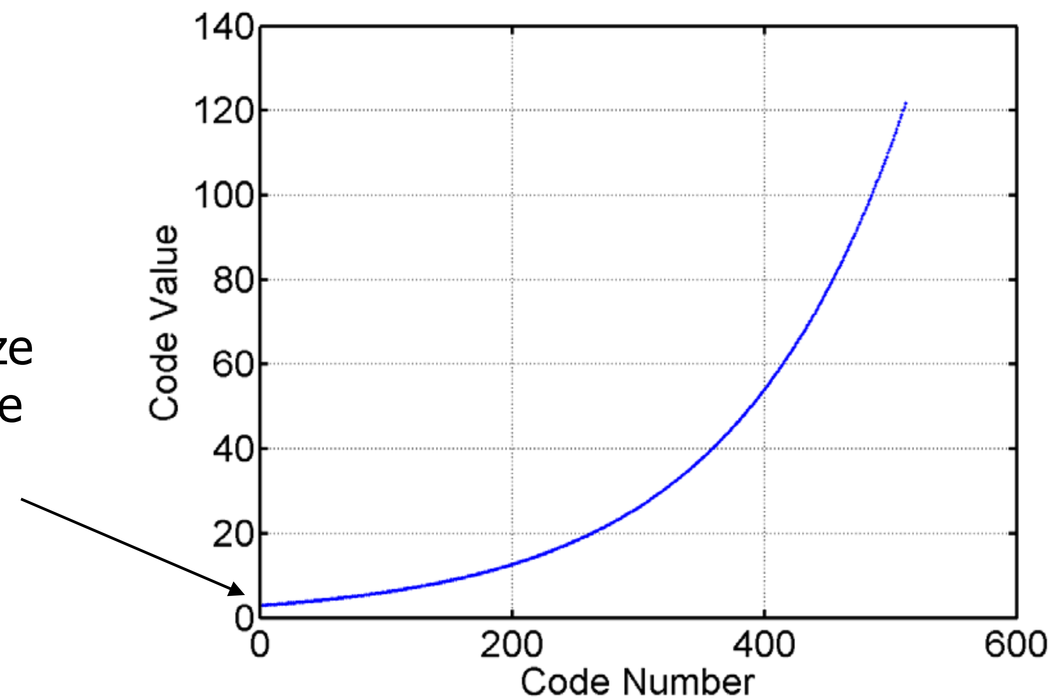
Objective to Subjective Correlation: S-T Region Size

- SCI / ISAS 2001 paper, "The Relationship Between Performance and Spatial-Temporal Region Size for Reduced-Reference, In-Service Video Quality Monitoring Systems"
- "iiis01.pdf" available at:
  - http://www.its.bldrdoc.gov/n3/video/documents.htm

Feature Quantization

- Logarithmic Quantizers Minimize Errors in Distance Measures – 9 bits is Sufficient

Don't Quantize Below Feature Perceptibility Threshold



Step 2: Distance Measures - Quality Parameters

- Convert Feature Streams To Quality Parameters
- Measure Perceptual Impairments For Each S-T Region
- Apply Perceptibility Threshold
  - Clip Features at Low End

Step 2: Distance Measures - Quality Parameters

- Apply Visual Masking Functions
  - (Out-In)/In, log (Out/In)
  - Separate Gain and Loss (i.e., Positive Part, Negative Part)
- Error Pooling (Over Space and Time)
  - Normally Worst Case (e.g., 5%)
  - Depends on Feature and S-T Region Size

Improved Error Pooling Lowbw (10 kbits/s) VQM

- Macroblocks (MB) (e.g., 3r x 3c x 2t)
  - Worse Case Processing Within MB
  - Localized S-T Impairments
- Generalized Minkowski
  - Removes Non-Linearity Before Fitting
  - $$Minkowski(P,R) = \sqrt[R]{\frac{1}{N}\sum_{i=1}^{N}\left|v_i\right|^P}$$

Step 3: Video Quality Mapping or Model (VQM)

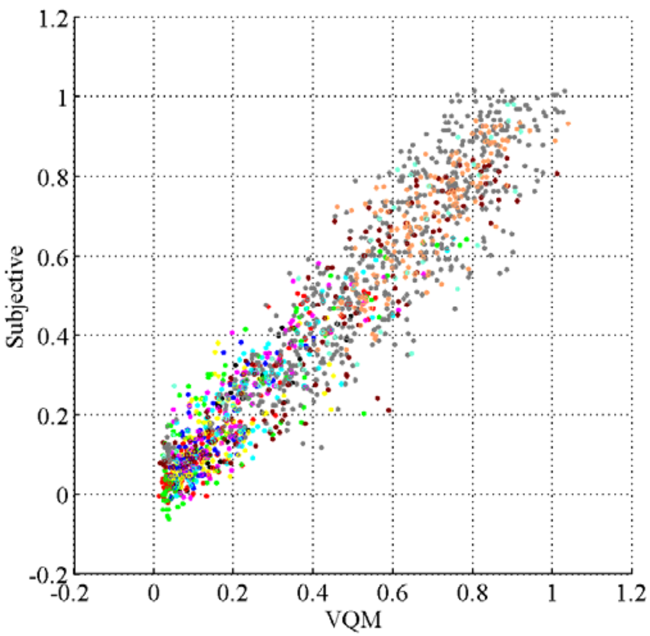
- Depends on Subjective Testing Methodology
  - e.g., DSCQS, SSCQE (See ITU-R Rec. 500)
  - Expert/Non-Expert Viewers
- Depends on Video Application
  - Television
  - Videoconferencing

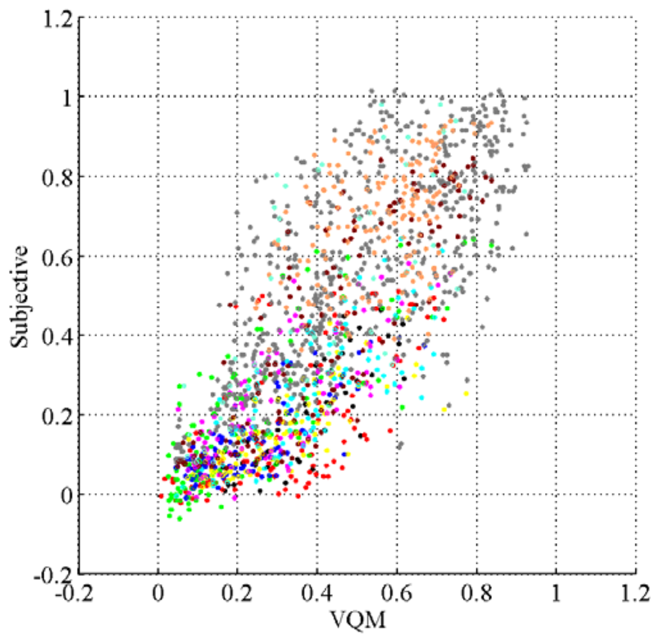
INLSA-Based VQMs

- ITS General VQM
  - 11 Subjective Data Sets (1563 Video Clips)
  - 7 Quality Parameters
- Lowbw (10 kbits/s) VQM
  - 19 Subjective Data Sets (2651 Video Clips)
  - 8 Quality Parameters
  - "vpqm05.pdf" available at:
    - http://www.its.bldrdoc.gov/n3/video/documents.htm

General VQM Performance – Training Data

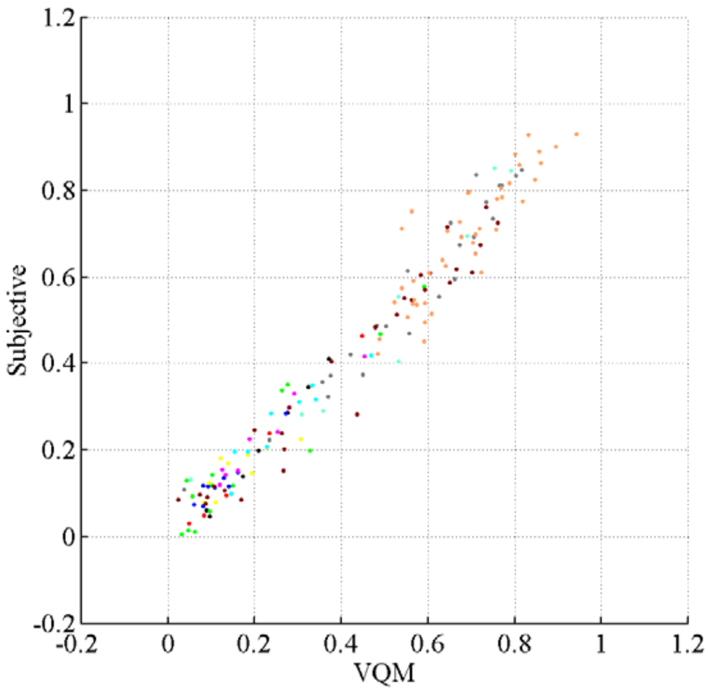
- Clip Performance (Scene x System)
  - ITS General VQM vs. PSNR VQM

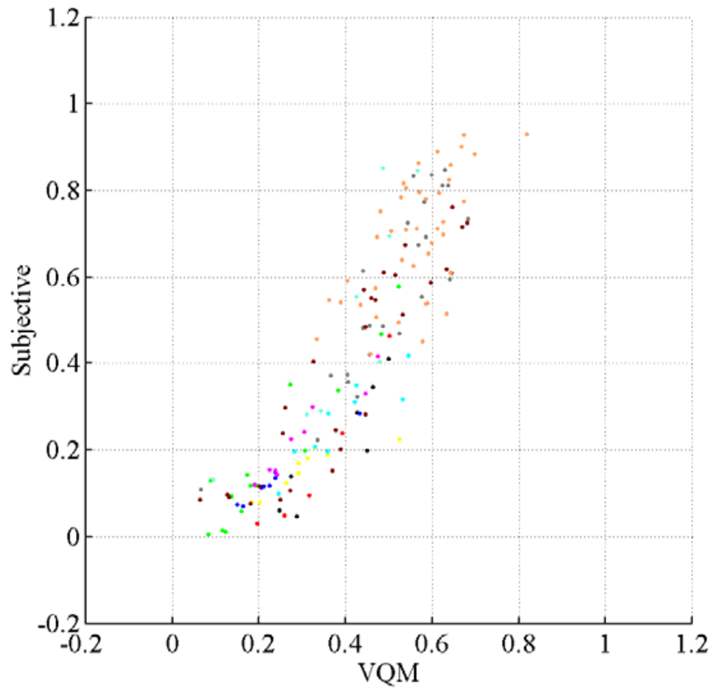




General VQM Performance - Training Data

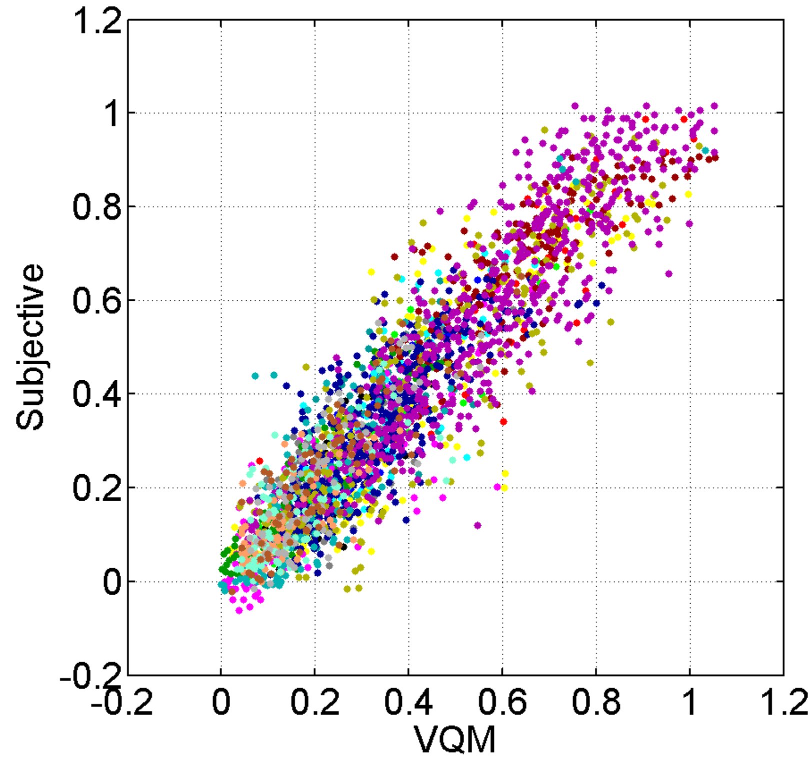
- Average Video System (HRC) Performance
  - ITS General Model vs. PSNR





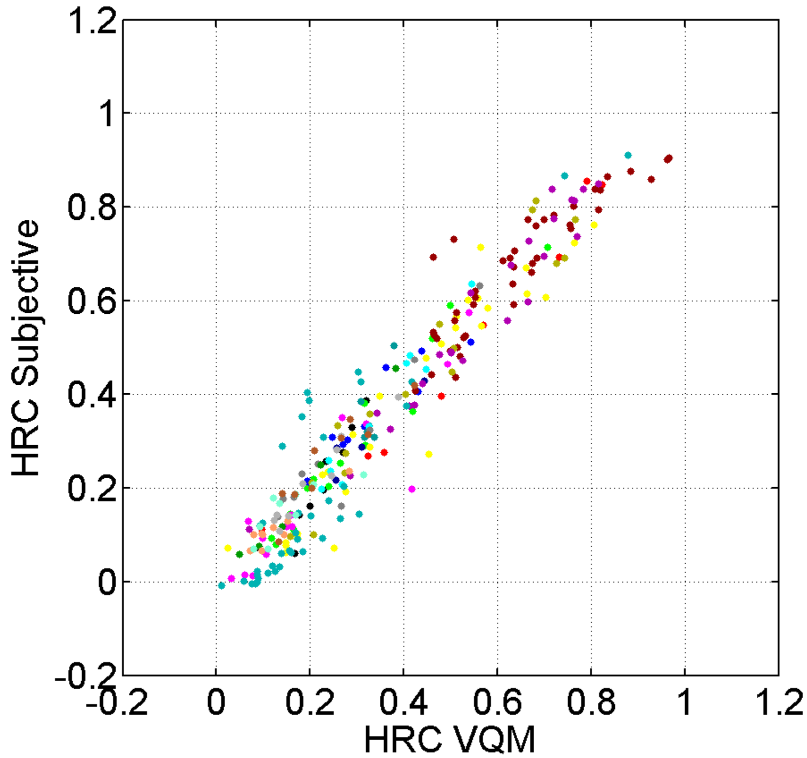
Lowbw (10 kbits/s) VQM Performance - Training Data

- Clip Performance
- ρ = .927
- RMS Error
  - 0.096
  - (0, 1) Scale



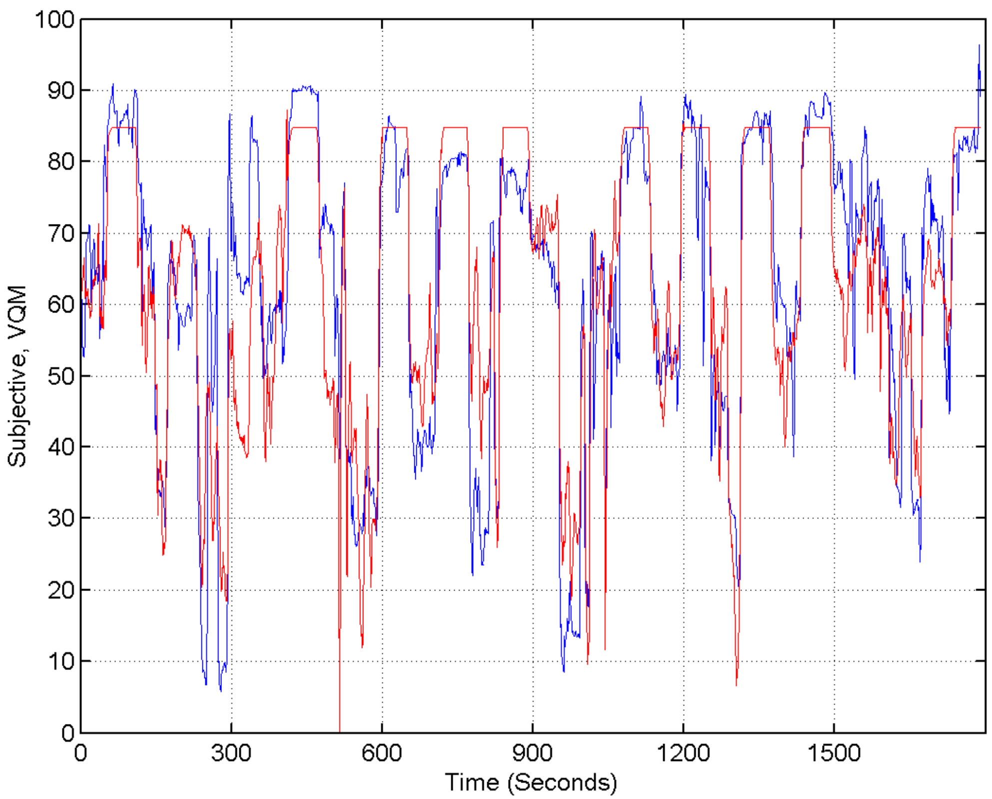
Lowbw (10 kbits/s) VQM Performance – Training Data

- HRC Performance
- ρ = .967
- RMS Error
  - 0.066
  - (0, 1) Scale



Continuous VQM Performance 10 kbits/s Prototype

- Single Stimulus Continuous Quality Evaluation (SSCQE)



Outline of Presentation

- New Measurements Required
- Subjective Video Performance
- Objective Video Performance
- Video Standardization Efforts
- Extensions of Technology
- Demonstration of Tools

Objective Metrics and Standards – 1990's Summary

- Four National Standards in Mid-1990s
  - ANSI T1.801.01-1995 (Test Scenes)
  - ANSI T1.801.02-1996 (Terms and Definitions)
  - ANSI T1.801.03-1996 (Objective Metrics)
    - Test Signals and Patterns Did Not Correlate with Subjective (Spinning Wheel, 3D Sine waves. etc.)
    - ITS Measurements Yielded 0.88 Correlation – Suitable for Quality Monitoring (Not System Comparison)
    - Spatial-Temporal Metrics Incorporated in ITU-T Rec. P.910 (Annex A) for Objectively Quantifying Scene Criticality
  - ANSI T1.801.04-1997 (Multimedia Delay, Synch, FR)
    - Later Standardized by ITU-T (P.931)

Objective Metrics and Standards – 1990's Summary

- Hardware Feasibility Demonstrated (Real-Time, In-Service, Perception-Based)
- Reduced Reference (RR) Measurement Paradigm Adopted by the ITU
  - ITU-T Rec. J.143

Video Quality Expert's Group (VQEG)

- Founded and Co-Chaired by ITS Staff (1997)
  - ITU-T SG 9 & 12, and ITU-R WP 6Q Experts
- ITS Manages
  - Website (www.vqeg.org)
  - Doc Server (ftp://ftp.its.bldrdoc.gov/dist/ituvidq/)
  - Video File Server (ftp://vqeg.its.bldrdoc.gov/)
- Mission – Validate Objective Video Quality Metrics & Forward Test Reports to ITU
  - Independent Lab Group (ILG) Conducts Tests
- ITU Writes and Approves Recommendations

Video Quality Expert's Group (VQEG) Full Reference TV (FRTV) Phase 1

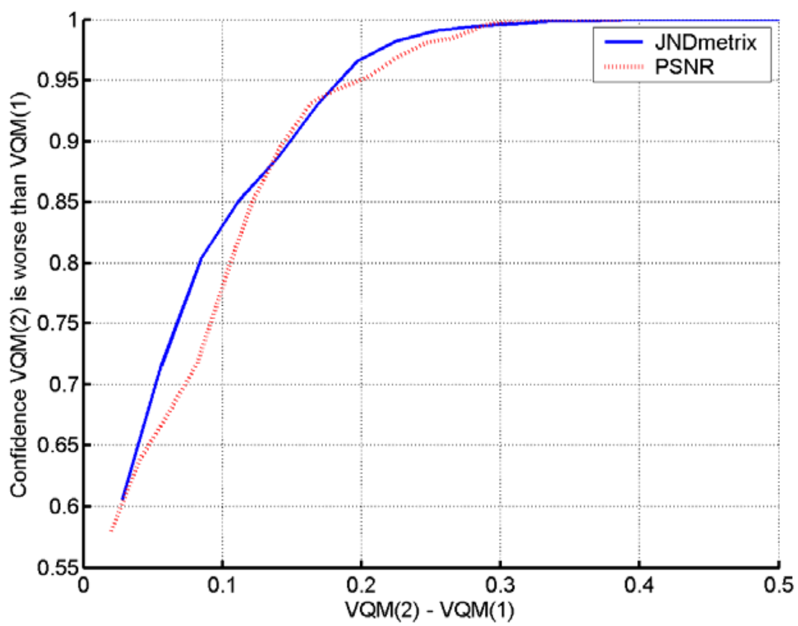
- From 1997 to 2000
- 8 Subjective Labs, 10 Objective Proponents
- 20 Source Sequences (split 525 & 625)
- 32 Video Systems
- No Clear "Winner"
  - ITU-T J.144 (March 2001)
  - VQEG Phase 1 Report

Committee T1 Technical Reports

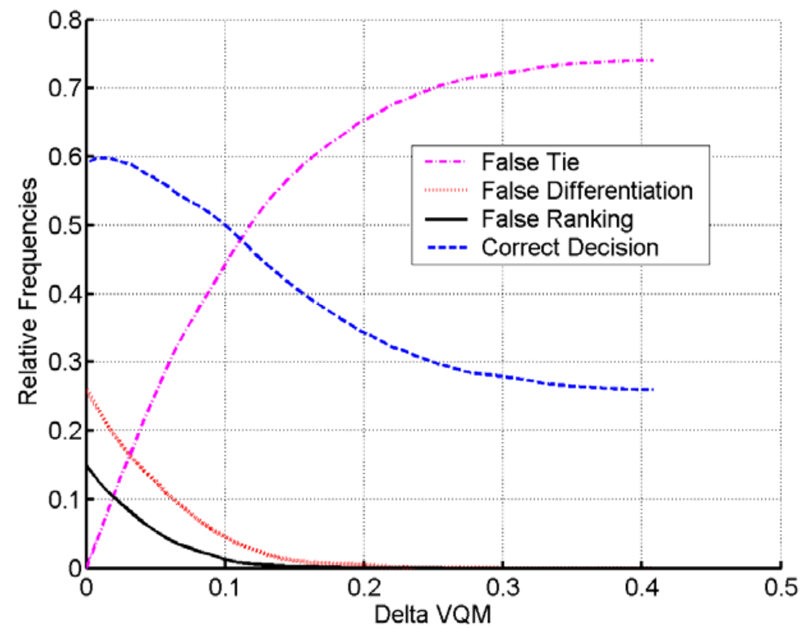
- From 2000 – 2002
- Jointly Developed by ITS and Sarnoff
  - VQEG Phase 1 Spin-off
- Five Committee T1 Technical Reports
  - T1.TR.72 - Specifying Accuracy and Cross Calibration of Video Quality Metrics
  - T1.TR.73, 74, 75, 77 - Calibration, PSNR, JND, Test Code and Data

Committee T1 Technical Reports

- Resolving Power & Classification Errors



Resolving Power



Classification Errors

Committee T1 Technical Reports

- Revised For Clarity
  - T1.TR.72-2003 (Approved Dec., 2003) "Methodological Framework for Specifying Accuracy and Cross-Calibration of Video Quality Metrics."
- International Recommendation
  - ITU-T Recommendation J.149 (Approved Mar., 2004) "Method for Specifying Accuracy and Cross-Calibration of Video Quality Metrics (VQM)."

Video Quality Expert's Group (VQEG) Full Reference TV (FRTV) Phase 2

- From 2000 to 2004
- 3 Subjective Labs, 8 Objective Proponents (2 Withdrew)
- 26 Source Sequences (split 525 & 625)
- 24 Video Systems
- Four Systems Were Recommended
  - ITU-T J.144 (March 2004)
  - ITU-R BT.1683
  - VQEG Phase 2 Report

ITS General VQM

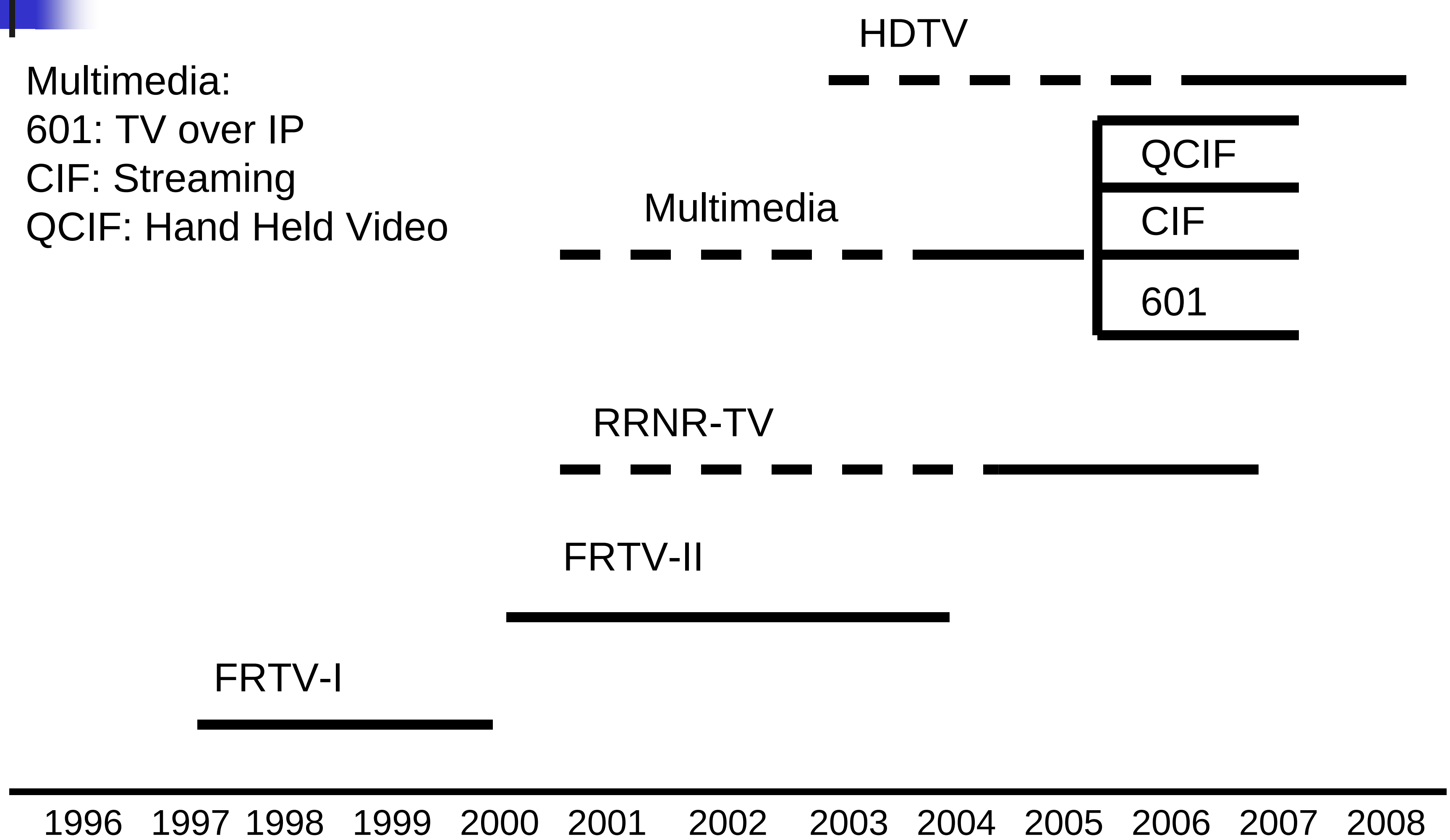
- Submitted to VQEG Phase II FRTV Tests
- Designed for Wide Range of Video Systems
  - Hence the name "General" VQM
- High Bandwidth RR VQM
  - Mbits/s of Reference Information
- Top Performing VQM in VQEG Tests
  - Only VQM in Top Performing Group (at 99% Level) for Both 525-line and 625-line TV Standards
  - Only VQM with an Average Correlation to Subjective Score > 0.9 (Over 525-line and 625-line Tests)
  - 0.94 Correlation to 525-line Subjective Test

ITS General VQM

- North American Standardization
  - ANSI T1.801.03-2003 (Approved Sept. 2003) “Digital Transport of One-Way Video Signals – Parameters for Objective Performance Assessment.”
  - Only Method Approved by ANSI for Video System Comparisons
  - Scope Does Not Include Transmission Errors

VQEG Timeline and Future Activities

Multimedia:
601: TV over IP
CIF: Streaming
QCIF: Hand Held Video



Joint Rapporteur Group on Multimedia Quality Assessment (JRG-MMQA)

- Recently Founded (2004)
  - Audio Experts from ITU-T SG12
  - Video Experts From ITU-T SG9
  - First Several Meetings – Revised Multimedia Test Plan

Outline of Presentation

- New Measurements Required
- Subjective Video Performance
- Objective Video Performance
- Video Standardization Efforts
- Extensions of Technology
- Demonstration of Tools

Multimedia to High Definition

Multimedia

- Different than TV
- Small, Low Resolution Screens (LCDs)
- Different Physics
  - Farther Viewing Distances (in terms of Picture Height)
  - Less Pixels per Degree
- Spatial Scaling Issues
- Different Applications
  - Access to Video Signals

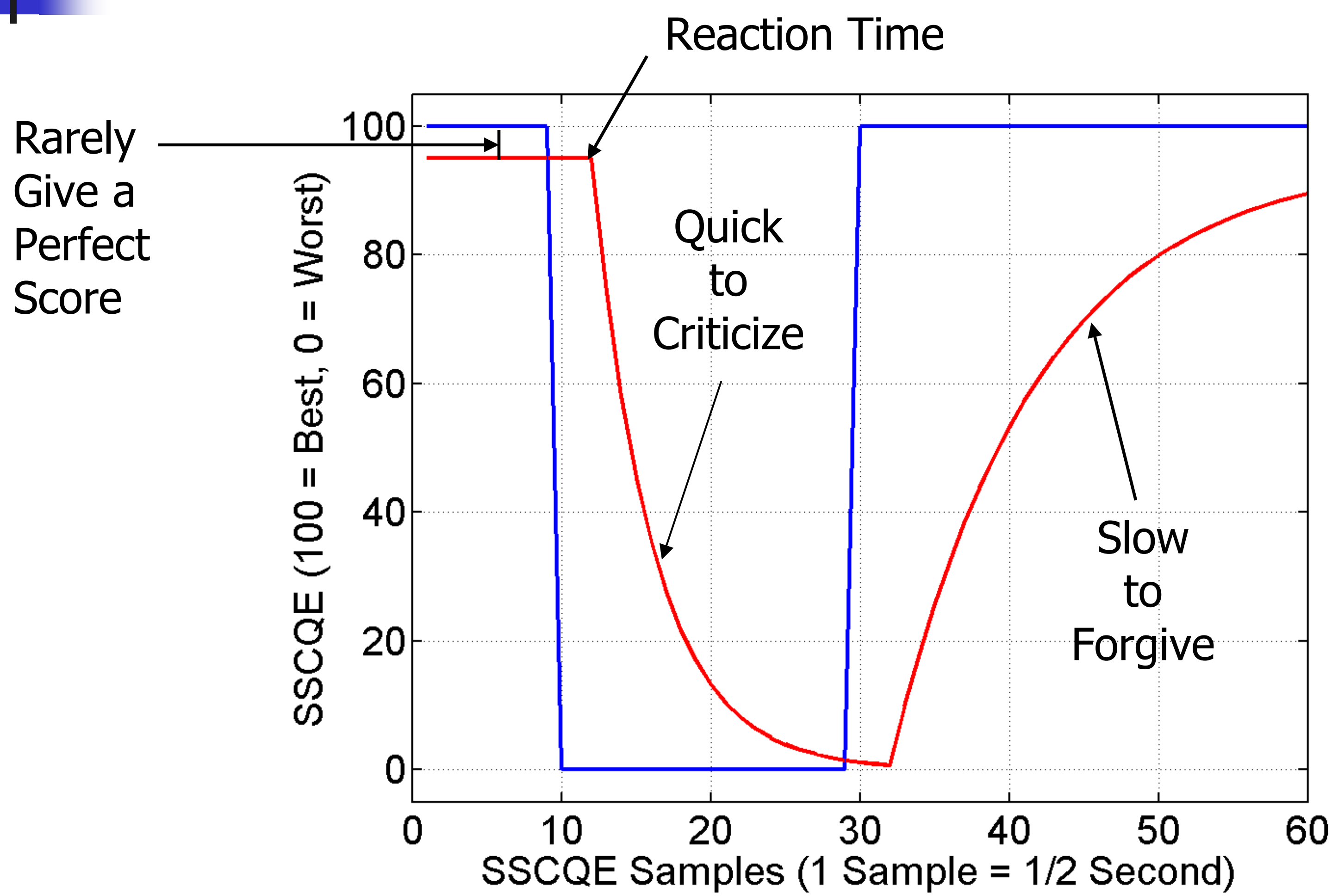
HDTV

- Different than TV
- Large High Resolution Screens
- Different Physics
  - Closer Viewing Distances (in terms of Picture Height)
  - Same Pixels Per Degree but Greater Viewing Angle
  - Subject Focus, Object/Scene Tracking Issues

Challenges / Questions

- Localized Errors (in space and/or time) can Have a Large Perceptual Impact
  - Very Small % of Video Stream
  - Thus, Difficult to Measure Robustly
- How Much Human Behavior Should We Really Model?
  - SSCQE Thought Experiment

SSCQE Thought Experiment



Outline of Presentation

- New Measurements Required
- Subjective Video Performance
- Objective Video Performance
- Video Standardization Efforts
- Extensions of Technology
- Demonstration of Tools

Laboratory VQM Tool Overview

- Designed for Bench Top Evaluation
  - Source and Destination Video Signals / files available at one PC
- UNIX (Batch) and PC (Interactive) Versions
  - PC Version Jointly Developed by ITS and Intel
  - Focus on PC Version Here

Laboratory VQM Tool Overview

- Two Software Flows
- Test Video Sequence (TVS)
  - Calibration Assisted by Special Test Pattern
  - U.S. Patent Pending
- Original (Source) and Processed (Destination) Library of Video Clips
  - Calibration Uses Scenes
  - All Clips Must be From Same HRC

Laboratory VQM Tool Overview

- Calibration Algorithms (Spatial, Temporal, Gain & Level Offset, Valid Region)
  - Includes Calibration Root Cause Analysis
- Five VQMs (General, Video Conferencing, TV, and Developers – Fast, PSNR)
  - Includes Impairment Root Cause Analysis
- Hierarchical Presentation of Results
  - HRC VQM -> Clip VQM -> Clip Parameter -> Parameter Time Histories -> Perceptual Features

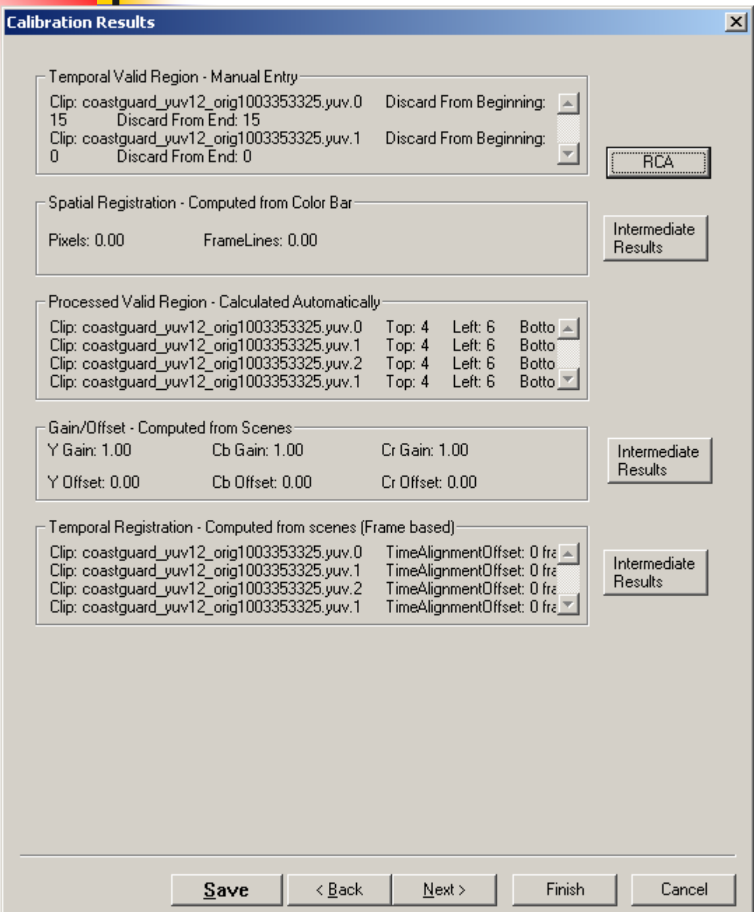
Laboratory VQM Tool Overview

- VQM Calculation is Multi-threaded
- VFW Frame Capture Included
- Video Conversion Tools Included
  - UYVY is Native Tool Format

Laboratory VQM Tool User's Manuals

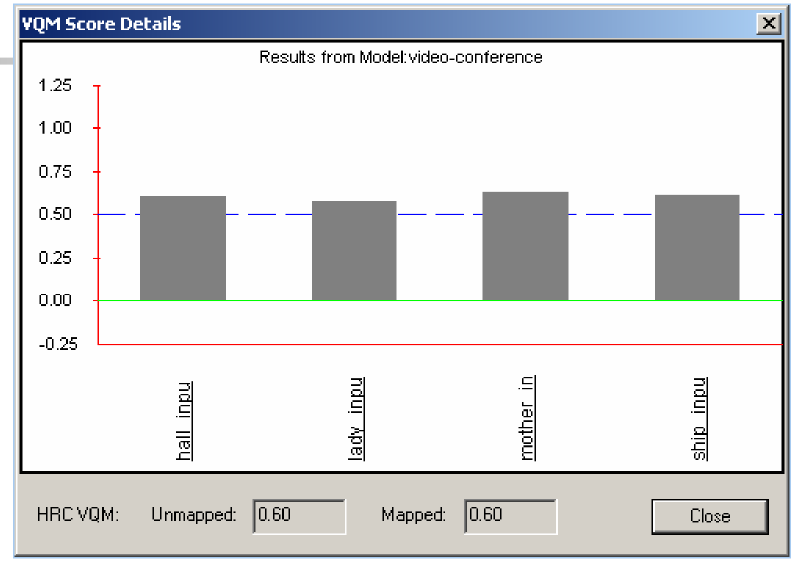
- UNIX – NTIA Handbook 02-01, "Video Quality Measurement User Manual," Feb., 2002.
- PC – Web PDF Document, "Video Quality Measurement PC User's Manual," Nov., 2002.

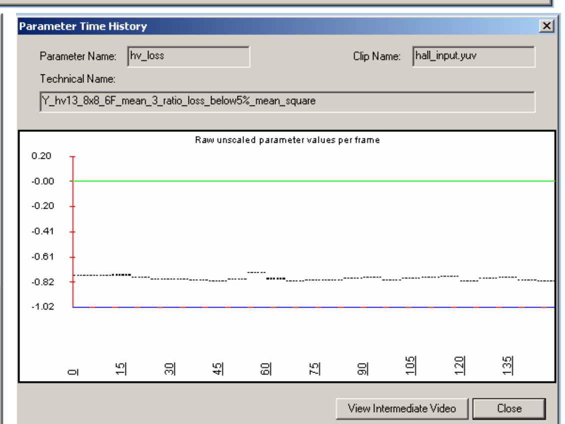
Laboratory VQM Tool



Calibration Results

Sample Screen Shots





VQM Results

In-Service VQM (IVQM) Tool Overview

- Developed by ITS (04 to 05)
- Perform End-to-End In-Service VQM (IVQM) Measurements
  - Requires Two PCs (Source and Destination)
  - RR Features Communicated over Internet
- Passive Monitoring of Video Stream
  - Attach Video Capture Probes at Two Points
  - Plot Measured Quality
- MATLAB-Based

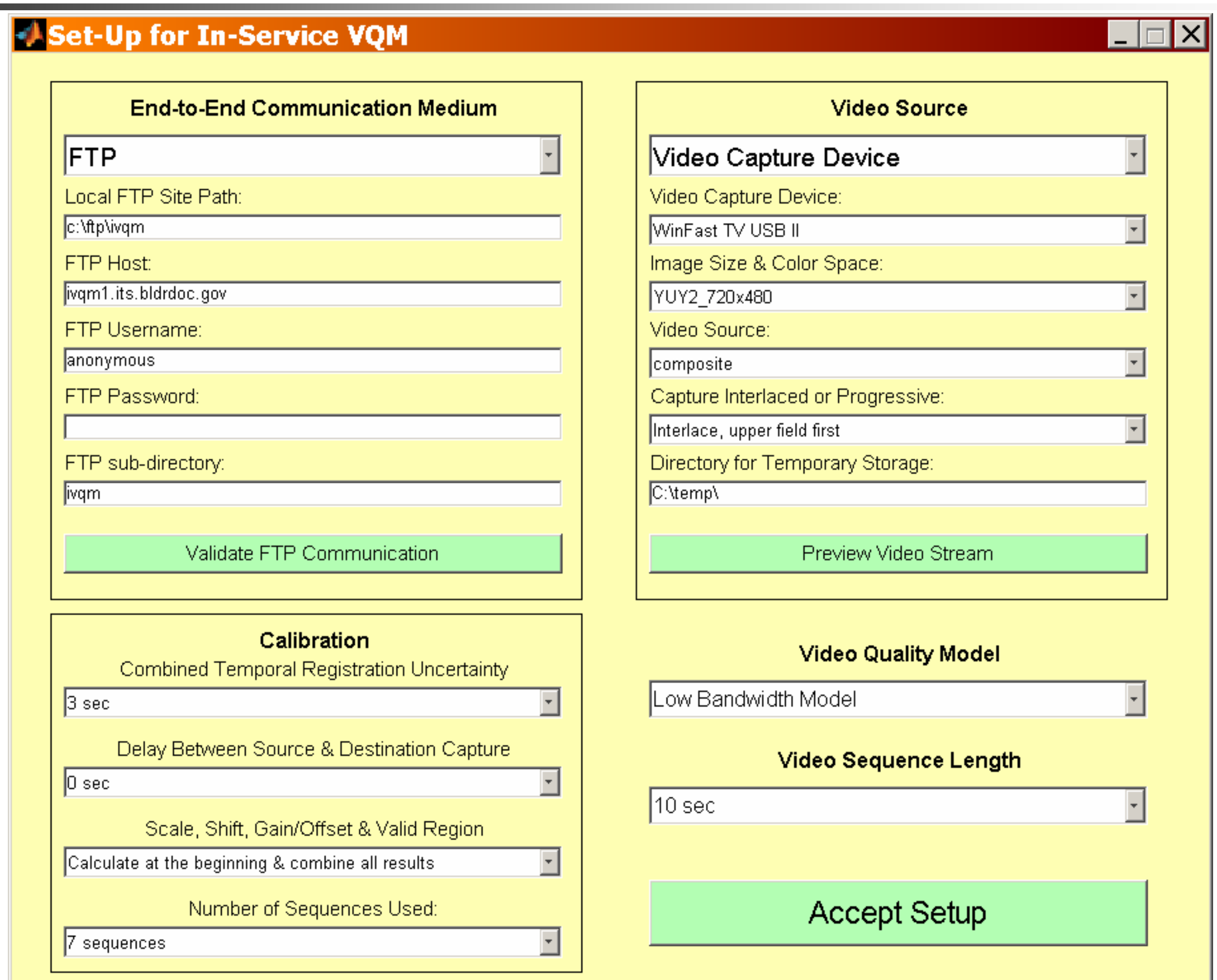
IVQM Tool Overview

- Supports Real Time Video Capture
  - e.g., USB2 Video Capture on Laptops
  - Large RAM Requirements
- Includes Full Calibration
  - e.g., Temporal Registration, Spatial Scaling, Spatial Registration, Valid Region, Gain / Level Offset
  - Video Delay with Synchronized Clocks
- Currently Implements 4 VQMs
  - General, Developer (Fast Version of General)
  - Lowbw, Fastlowbw

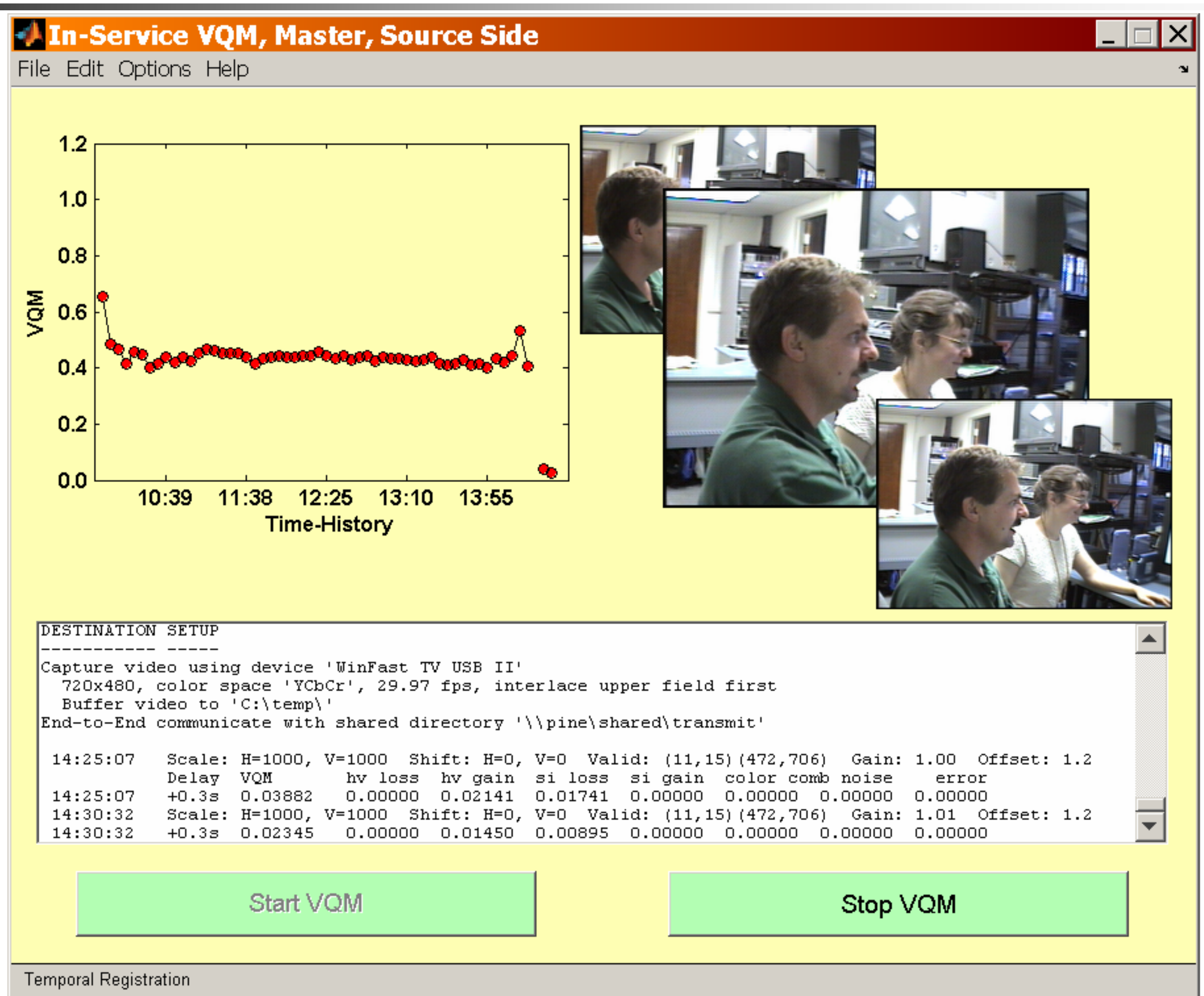
IVQM Tool Overview

- Non-Real Time Measurements
  - Depends on VQM and Image Resolution
- Captures Sequence -> Analyzes Sequence -> Captures Next Sequence
- User's Manual
  - NTIA Handbook, "In-Service Video Quality Metric (IVQM) User's Manual", Dec, 2005.

IVQM Tool Setup Screen



IVQM Tool Monitoring Screen



New Measurement Tools (Available Shortly)

- Command-line VQM (CVQM)
  - Simple Window's Command Line Interface
- Batch VQM (BVQM)
  - GUI Interface
  - Greatly Simplify Batch Processing on PC
- Dynamically Linked Library (DLL)
  - Standard Interface for Third Party Applications