

# Objective Estimation of Video and Speech Quality to Support Network QoS Efforts



2<sup>nd</sup> Internet2/DoE Quality of Service Workshop  
February 10, 2000, Houston, Texas

Stephen Voran & Stephen Wolf  
Institute for Telecommunication Sciences  
Boulder, Colorado

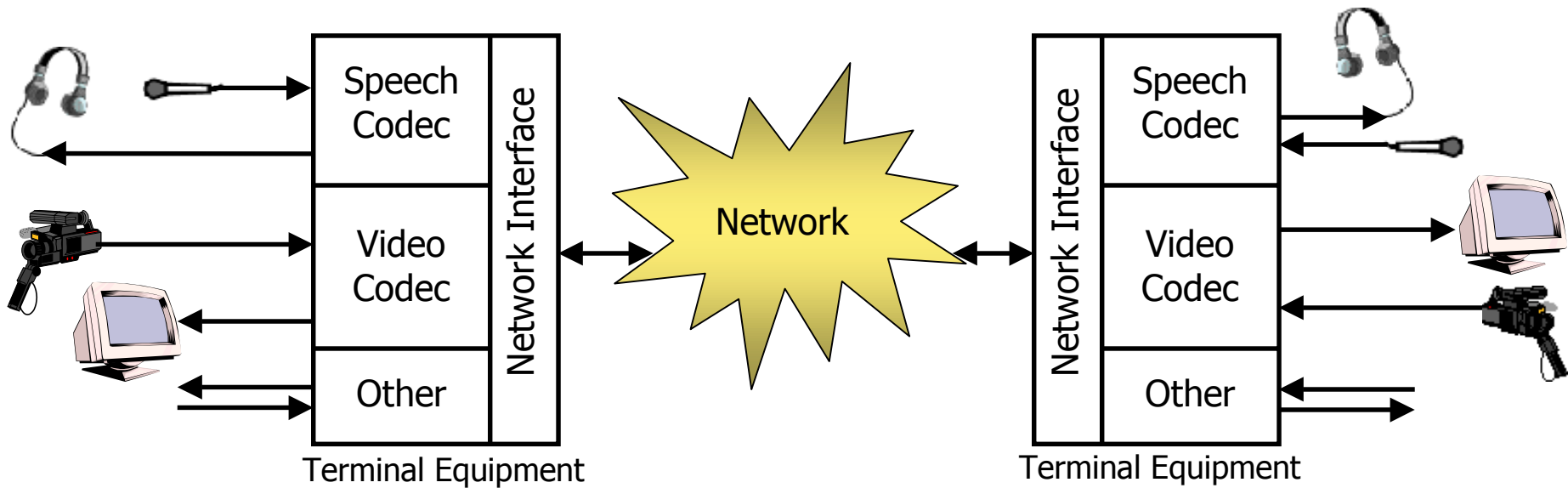


# Motivation

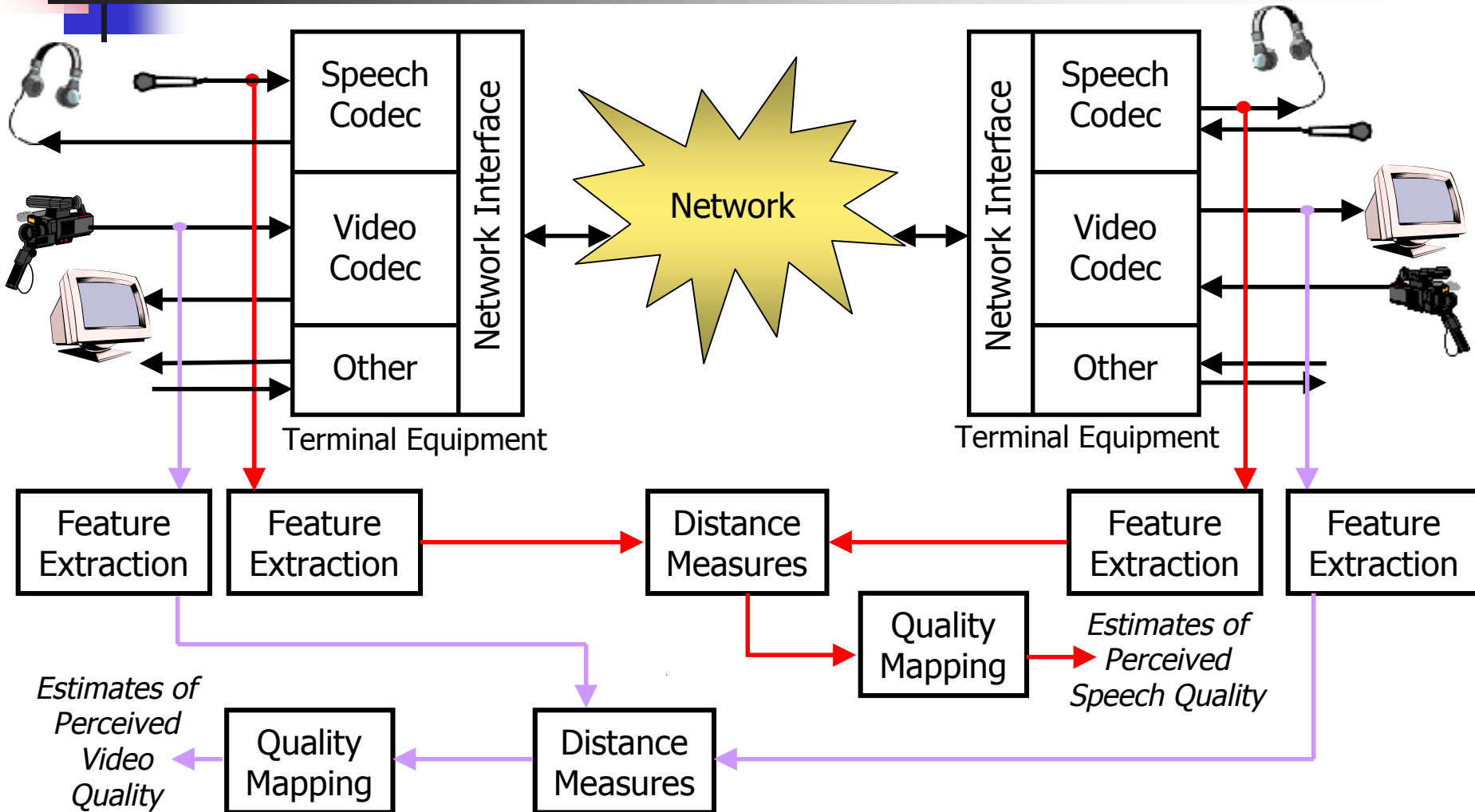
---

- User-perceived QoS is driven by many factors, including
  - Perceived quality of delivered video
  - Perceived quality of delivered speech
- From the user point of view, these are two important components of QoS
- From the network point of view, these are demanding applications

# Example Multimedia Communication System



# Quality Estimation for Multimedia Communication System (In Service)



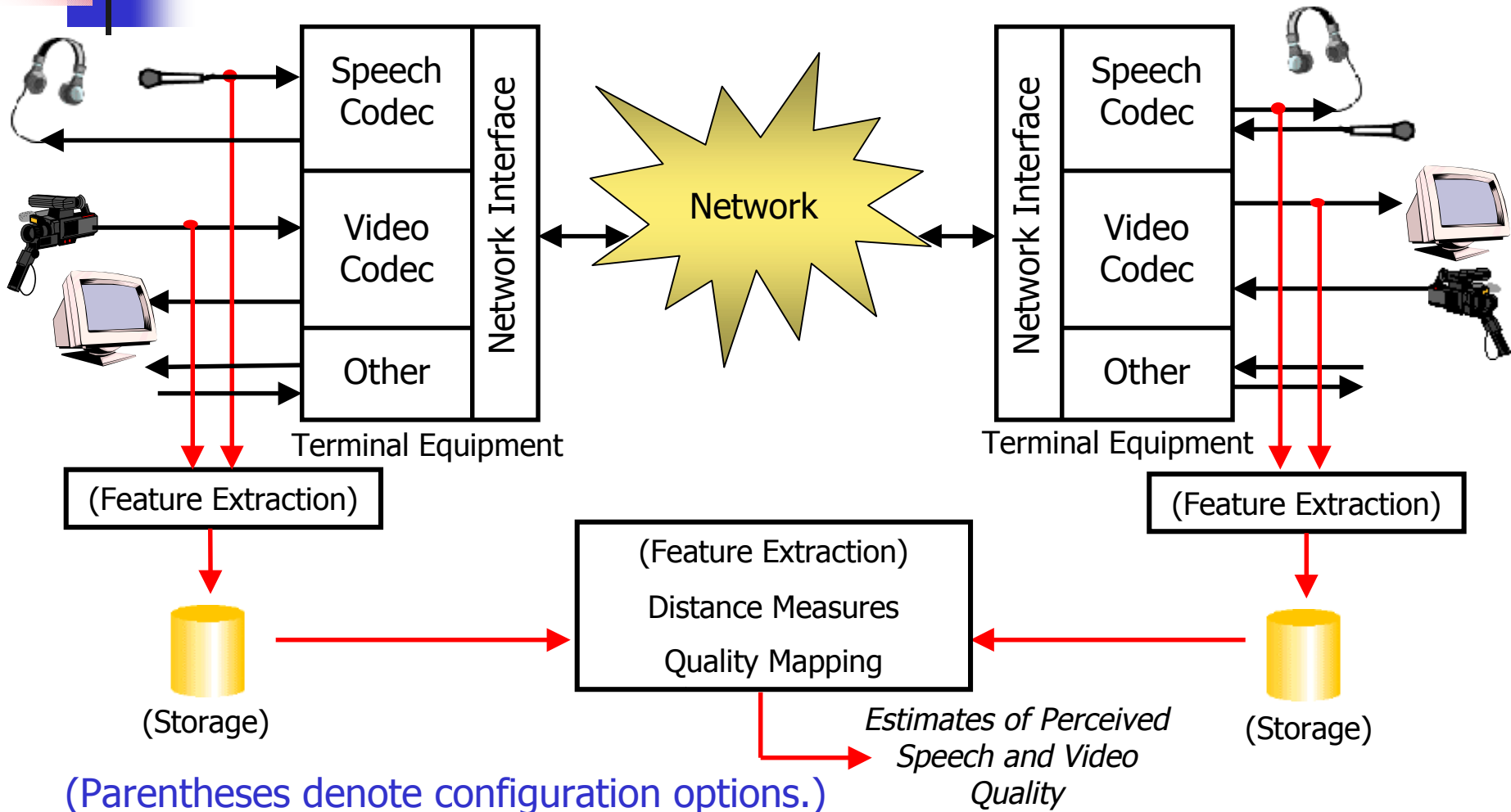


# Quality-Estimation Components

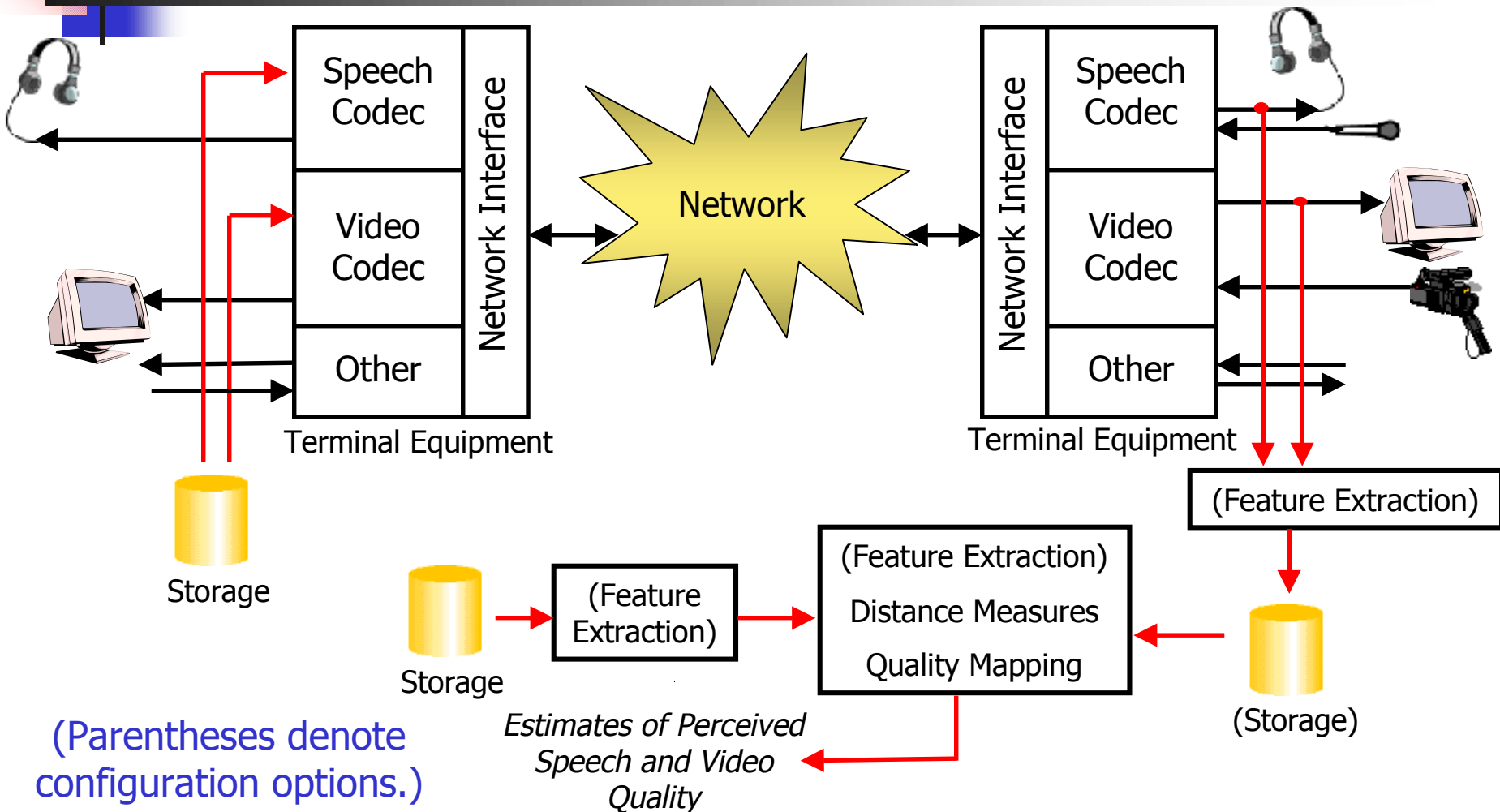
---

- Feature Extraction:
  - Extract perceptually-relevant, quality-related attributes of signals
  - Also called perceptual transformations
  - $.0025 < (\text{feature bit-rate}/\text{signal bit-rate}) < 1$
- Distance Measures:
  - Calculate perceptually-relevant distance between two streams of features
- Quality Mappings:
  - Relate perceptual distances to estimates of perceived quality

# Quality-Estimation Configurations (In service)



# Quality-Estimation Configurations (Out of service)





# Speech Quality Estimation Details

---

- Feature extraction or perceptual transformation:
  - Transformation from Hz scale to critical band or Bark scale
  - Transformation from signal amplitude to an approximation of loudness
- Distance measure:
  - Measures and normalizes out spectral errors at successively smaller time and frequency scales





# Speech Quality Estimation Status

---

- Measuring Normalizing Block (MNB) algorithms provide reliable quality estimates for 4-kHz bandwidth speech, through waveform and non-waveform codecs with bit-errors and frame erasures and constant delay
- Continuing challenges include:
  - Variable transmission delay
  - Wider bandwidth speech
  - More general audio signals



# Video Quality Estimation Details

---

- Feature Extraction
  - Transform video using perceptual filters
    - Spatial information (magnitude, direction)
    - Temporal information
    - Chroma information
  - Extract summary statistics from S-T regions
    - Standard deviation, mean
    - Region size determines compression factor
    - 8 lines x 8 pixels x 6 frames



# Video Quality Estimation Details

---

- Distance Measure
  - Apply perceptibility threshold
  - Apply visual masking function
    - Loss:  $\{out-in\}/in$
    - Gain:  $\log \{out/in\}$
  - Pool error over space (e.g., worst 5%)
  - Pool error over time

# Spatial Information Loss (e.g., Blurring)



Spatial  
Filter



Spatial  
Filter



# Spatial Information Gain (e.g., Tiling)



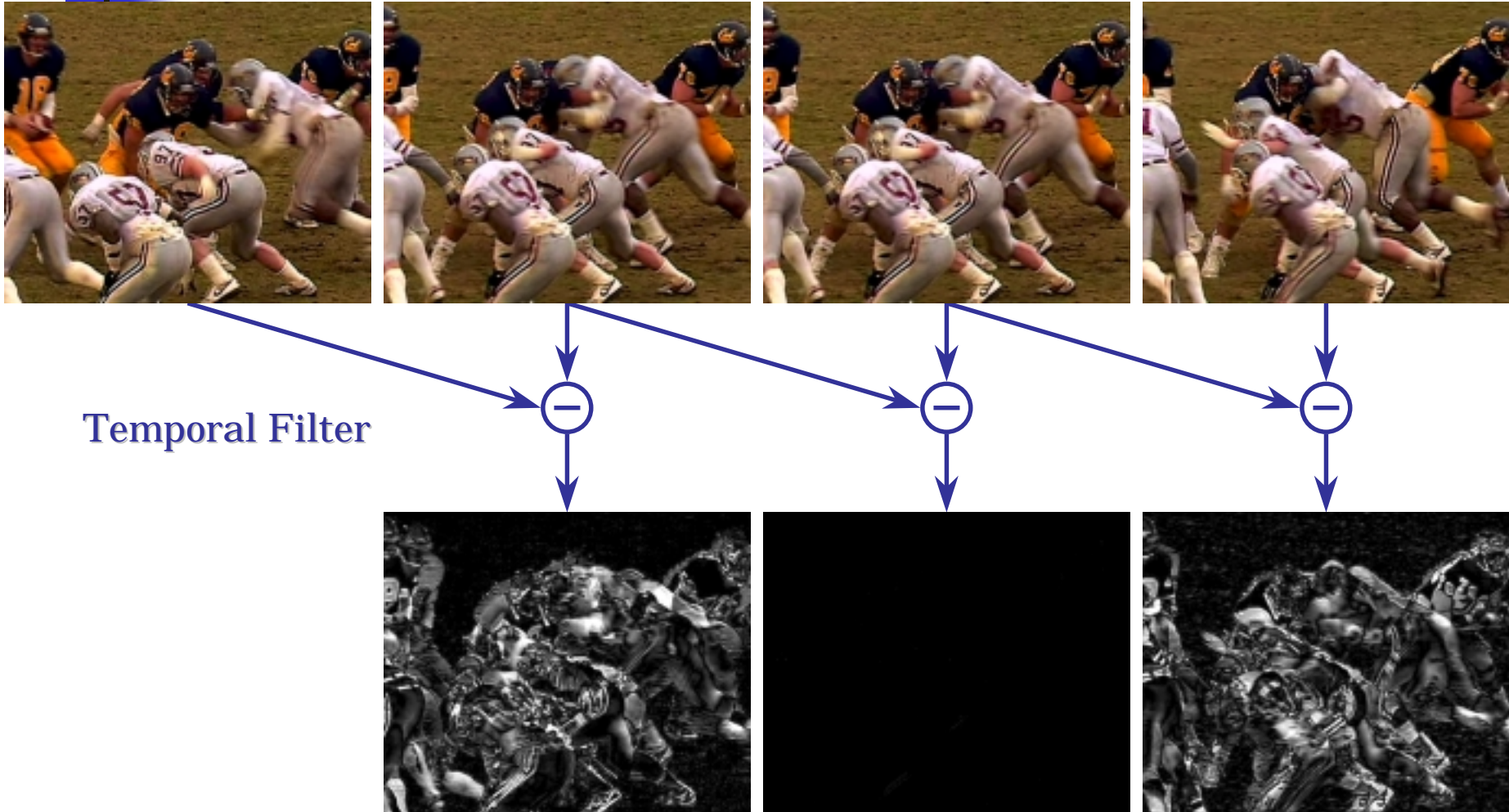
Spatial  
Filter



Spatial  
Filter



# Temporal Information Loss (e.g., Dropped Frames)



# Temporal Information Gain (e.g., Error Blocks)

Frame n

Frame n-1





# Video Quality Estimation Status

---

- Wide range of scenes, systems, bit-rates
  - Spatial detail, motion, contrast, brightness, noise, saturation, hue
  - VTC, MPEG-1/2/4, DS3, analog, digital transmission errors, multi-generation
- Continuing challenges include
  - Variable transmission delay systems
  - Continuous quality evaluation
  - Impairment visibility for some scenes





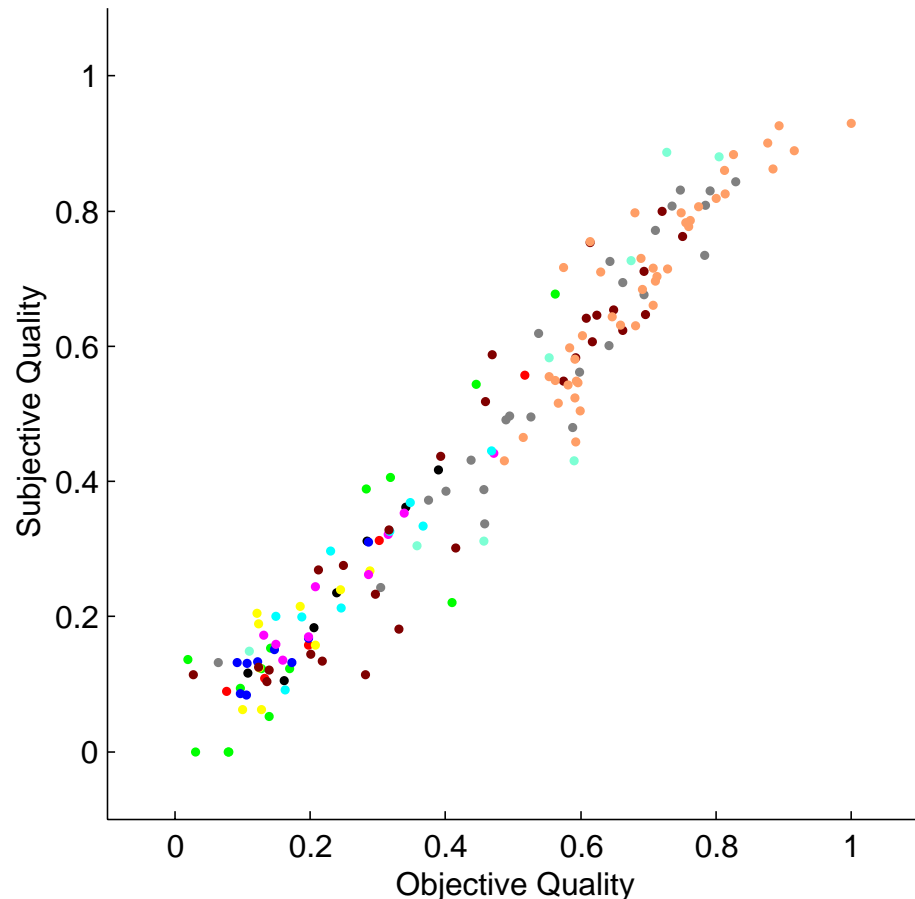
# Typical Performance of Estimators

---

- Measured through correlation to results of controlled, formal, subjective listening and viewing tests
- Per System Correlation Coefficient
  - $.70 \leq \rho^2 \leq .98$  (speech)
  - $.74 \leq \rho^2 \leq .98$  (video)

# Example Scatter Plot for Video Subjective vs. Objective

- Eleven Data Sets
  - 164 systems
  - $\rho^2 = .95$
  - Zero = no impairment
  - One = max impairment





# Standardization

---

- Published standards
  - Video: ANSI T1.801.03-1996 and ITU-T P.910, Section 5.3 and Annex A, 1996
  - Speech: ANSI T1.518-1998 and ITU-T P.861 Appendix II, 1998
- Extensions, refinements and additional standardization efforts continue



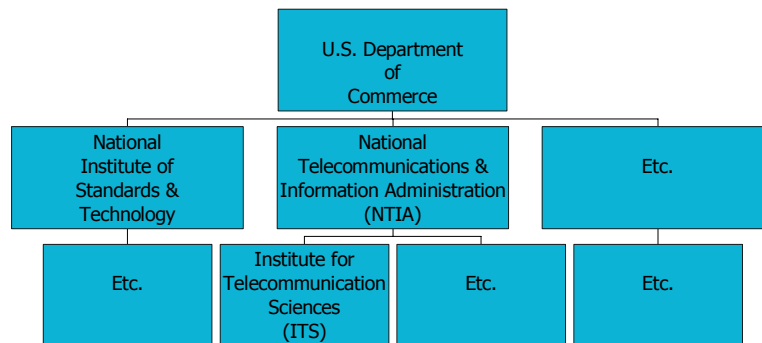
# Conclusions

---

- Perception-based objective estimators of video and speech quality have been derived and verified
- Might they support network QoS efforts?
- We welcome opportunities for collaborative experimentation



# Institute for Telecommunication Sciences



ITS is the chief research and engineering arm of NTIA. ITS mission includes:

- Promotion of advanced telecommunications and information infrastructure development in the US
- Facilitation of more efficient and effective use of radio spectrum
- Enhancement of domestic competitiveness and foreign trade opportunities of US telecommunications firms



# For more details

---

- [www.its.bldrdoc.gov](http://www.its.bldrdoc.gov)
- S. Voran, "Objective Estimation of Perceived Speech Quality, Part I: Development of the Measuring Normalizing Block Technique," IEEE Transactions on Speech and Audio Processing, July 1999
- S. Voran, "Objective Estimation of Perceived Speech Quality, Part II: Evaluation of the Measuring Normalizing Block Technique," IEEE Transactions on Speech and Audio Processing, July 1999
- G.W. Cermak, S. Wolf, E.P. Tweedy, M.H. Pinson, and A.A. Webster, "Validating Objective Measures of MPEG Video Quality," SMPTE Journal, April 1998
- S. Wolf, "Measuring the End-to-End Performance of Digital Video Systems," IEEE Transactions on Broadcasting, September 1997