

Perception of Extremely Low-Rate Images & Video: Psychophysical Evaluations and Analysis

Sheila S. Hemami

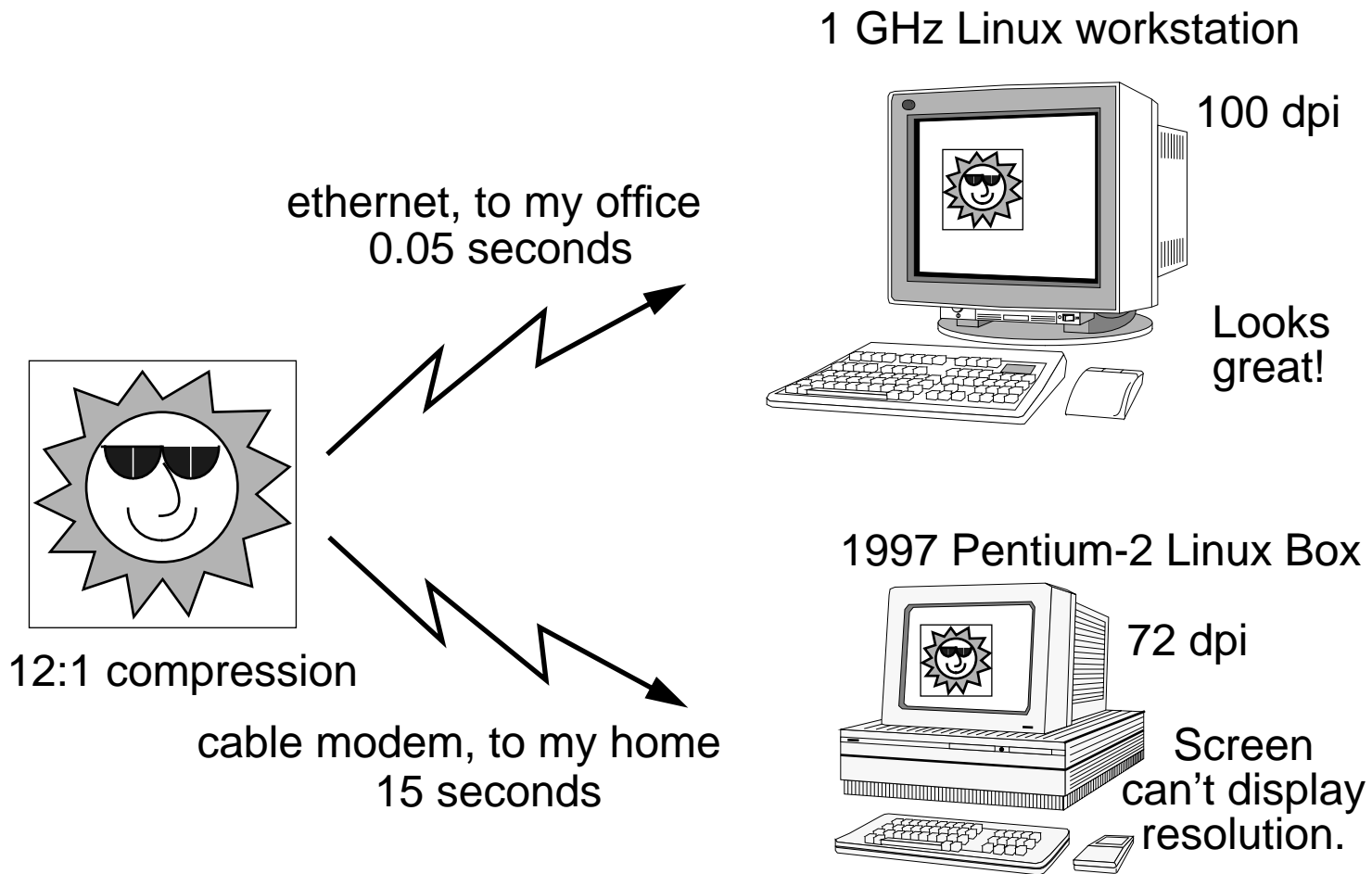
School of Electrical Engineering
Cornell University
January 2001

Visual Communications Lab
<http://foulard.ee.cornell.edu>

Why Use Psychophysics?

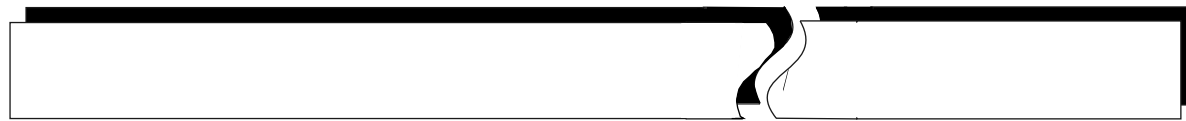
- At low bit rates, “every bit counts.”
- * *Psychophysics* — a branch of psychology concerned with the effect of physical processes (e.g. an intensity of stimulation) on the mental processes of an organism.
- We need to characterize the importance of the signal and distortion on the perceived quality of an image. We do not yet have a mathematical model for perceived quality.

Image Transmission & Low-Bandwidth



Scalable Coding Decodes at All Rates, but....

One stream describes the image at all bit rates....but the low-rate images have visible artifacts...



0.01 bpp



0.3 bpp



Visual Optimization?

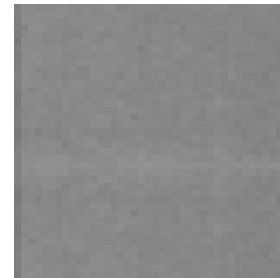
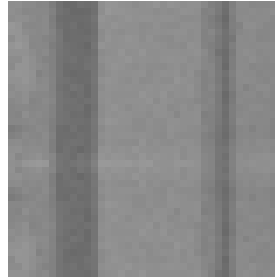
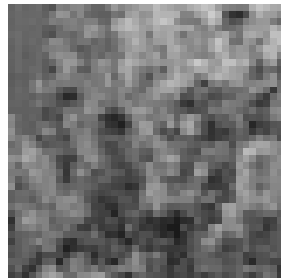
- If the image transmission is stopped early, the decoded image exhibits obvious artifacts.
- Can we maximize the visual quality *regardless of the bit rate*?
- This requires an understanding of perception of distortion that exceeds visibility thresholds, i.e., the perception of *suprathreshold* distortions.
- Current incorporation of HVS characteristics into compression algorithms are for *visually lossless compression* (i.e., looks perfect).

Outline

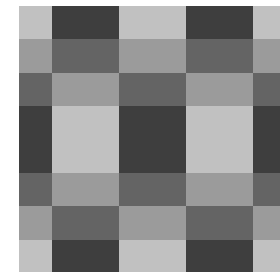
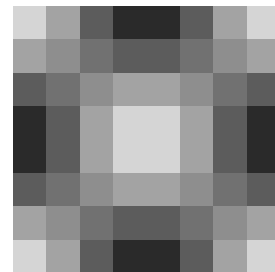
- The human visual system and still imaging
- Wavelet coding of images
- Quantifying sensitivities to supra-threshold quantization noise in complex stimuli
- Analysis — some new results on how we see in the suprathreshold regime
- Applications to compression
- Some preliminary video results

The Human Visual System, for our purposes

- Activity sensitivities

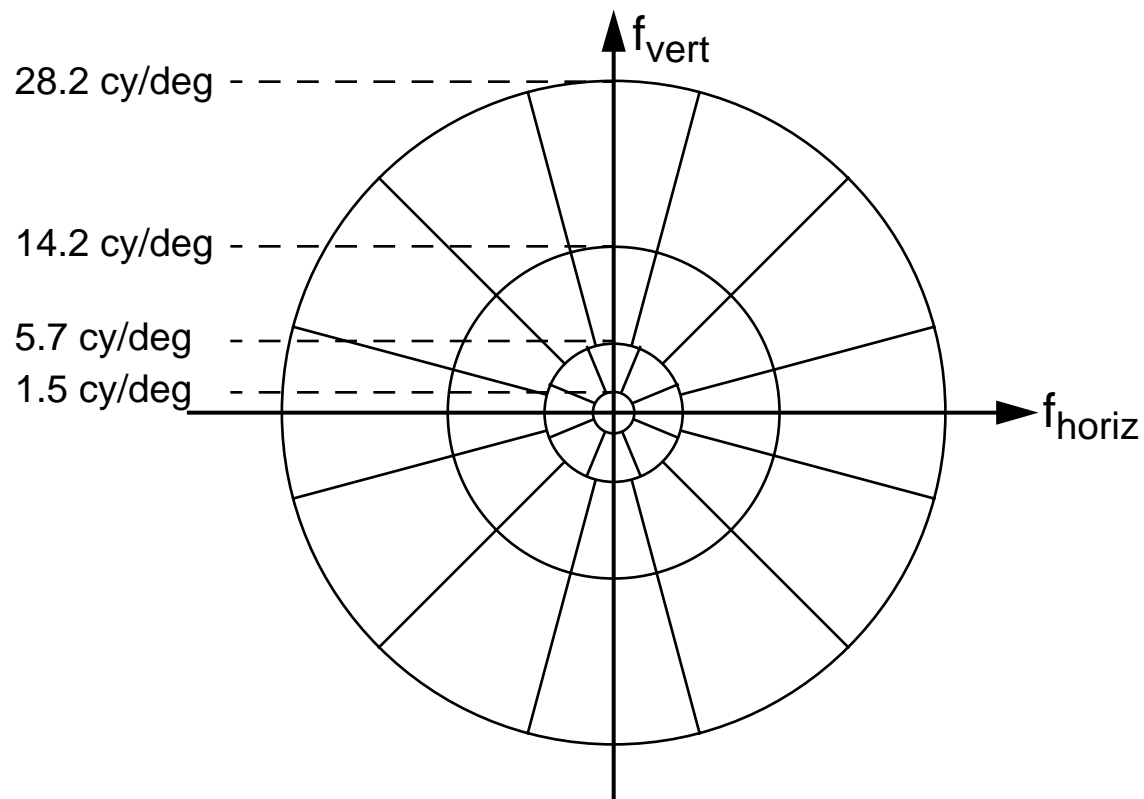


- Frequency sensitivity

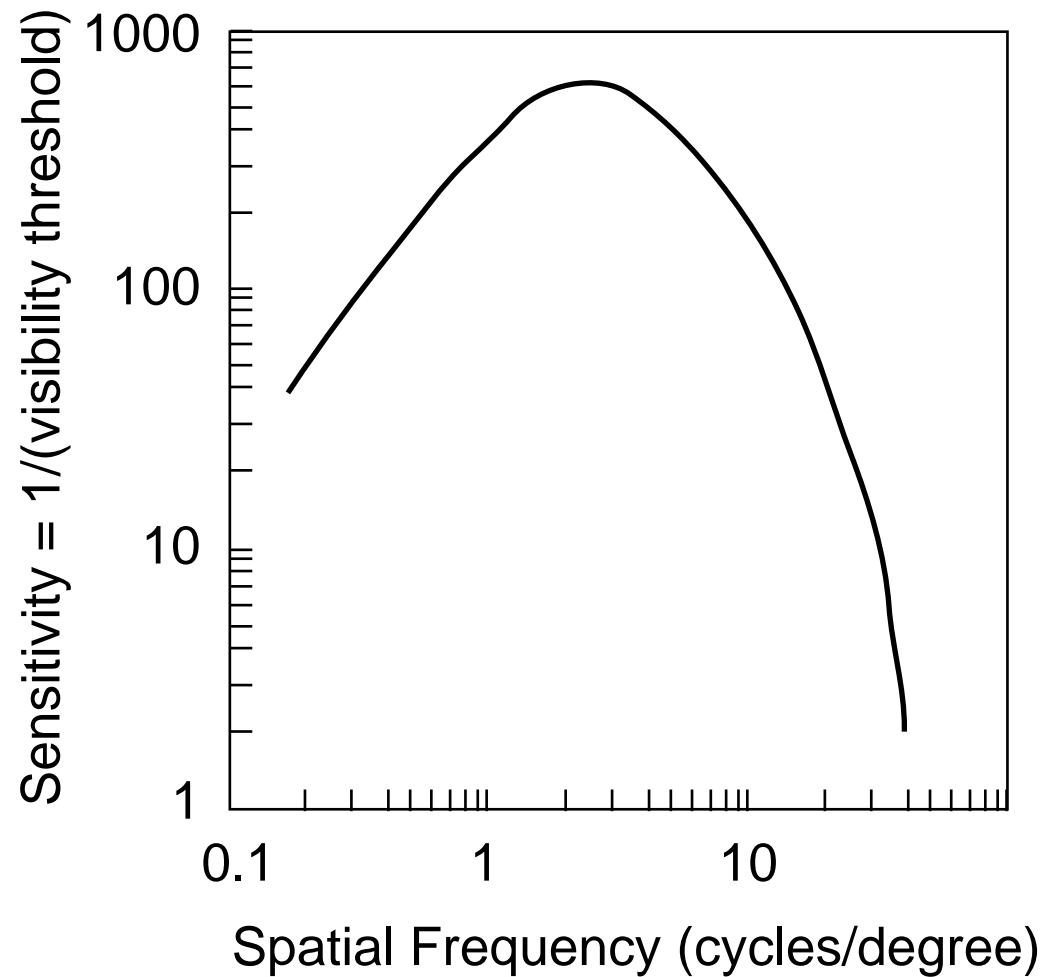


Multi-Channel Model of the HVS

The HVS consists of numerous channels, each tuned to a specific spatial frequency and orientation.

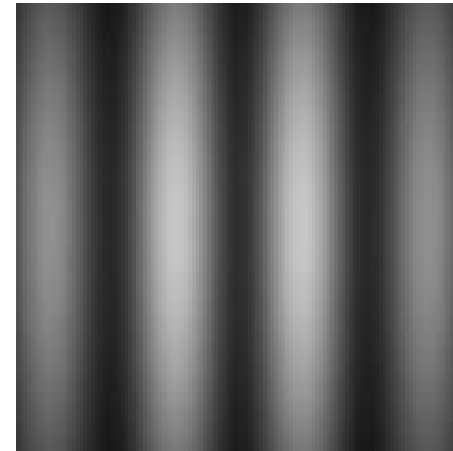


Human Contrast Sensitivity Function (CSF)



Experimental Determination of the CSF

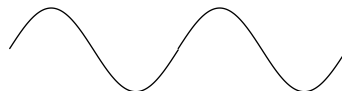
- A *simple grating* is a function of frequency and orientation.
- Observer increases the amplitude until the stimulus is visible.



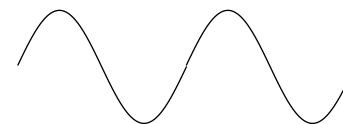
increasing amplitude



not visible



not visible



visible!

Visibility Threshold and Contrast Sensitivity

- The *visibility threshold* (VT) is the *contrast* at the point when the stimulus just becomes visible.
- For gratings with maximum/minimum luminances $L_{max/min}$,

$$\text{contrast} \equiv \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$$

(NOTE: there are many definitions of contrast.)

- Contrast sensitivity (CS) is given by

$$CS = 1 / VT$$

Some Comments on the CSF

- The CSF is not orientation specific.
 - But the HVS has channels tuned to different orientations as well as frequencies.
- The CSF is for simple gratings only.
 - But natural images contain many frequencies.
- The CSF represents *subthreshold* perception.
 - It is most accurate when predicting the response to stimuli just below the VT.

Visibility Thresholds in Natural Images

Visibility threshold (VT) — the point at which differences between a control and a sample image are visible to an observer.

VT is a function of the stimulus:

1. f (frequency sensitivity) — CSF
2. f (luminance masking) —

Weber's law: $VT \propto$ background luminance

The contrast of the stimuli must increase with the background luminance in order to stay equally visible.

3. f (contrast masking) (also called texture masking)
VTs for one image component are affected by the presence of another image component.

* The combination of luminance masking and contrast masking is referred to as *spatial masking*.

Spatial Masking Example: Noise Suppression

- Add Gaussian noise to 2 different regions and ask observers which image looks better.



- *Detailed* regions mask the most noise, followed by *edge* and then *smooth* regions.

Spatial Masking Example: In Natural Images

- Compare VTs of two stimuli:
 - Stimulus #1: Quantize a single wavelet subband in a wavelet-coded image; leave others unchanged.
 - Stimulus #2: Synthesize noise in a single wavelet subband; all other coefficients are 0.
- Results:

$$VT_1 \approx 2.5 VT_2$$

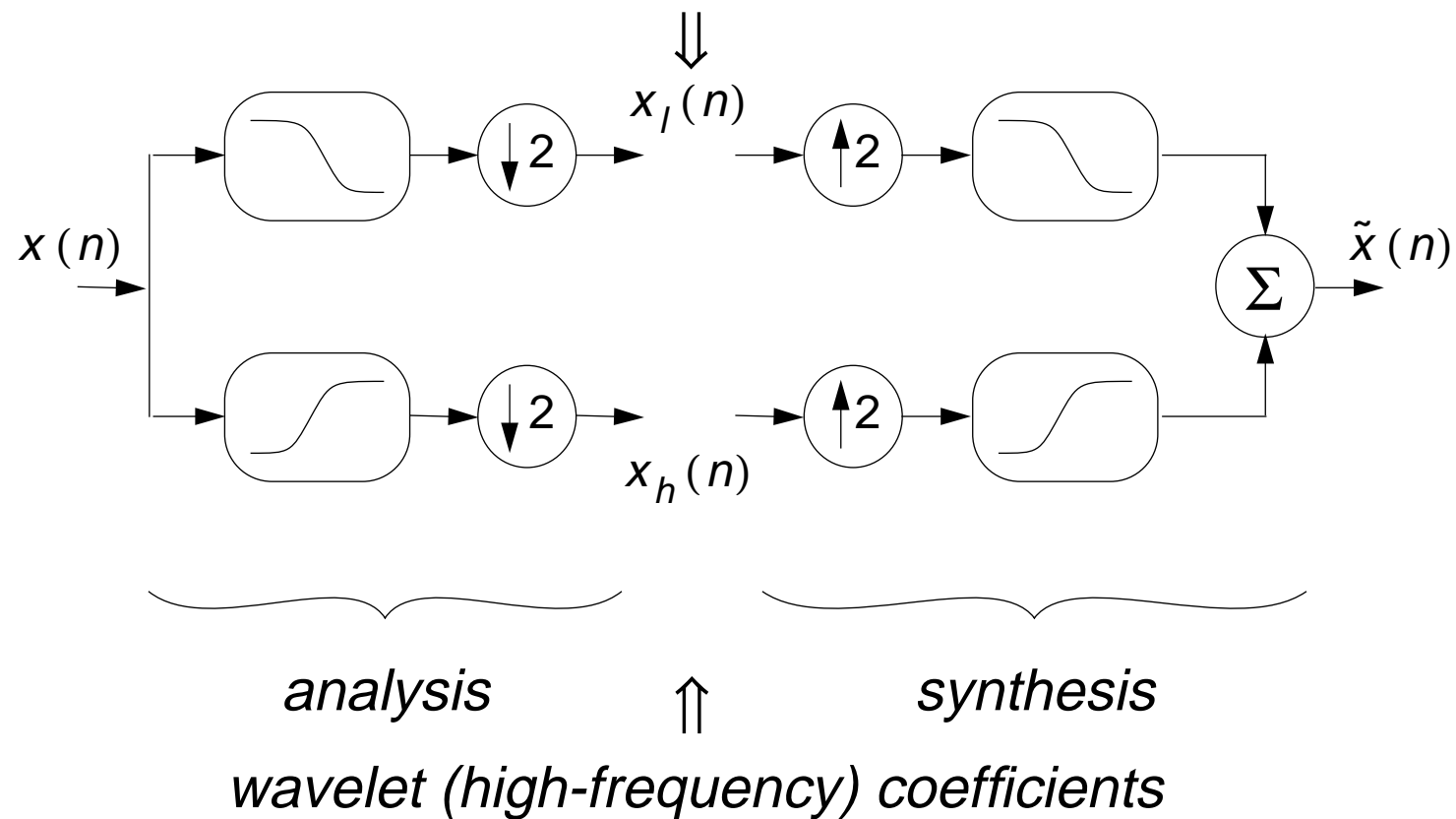
With the additional image signal present, the VT is elevated by a factor of approximately 2.5.

Quantifying Spatial Masking?

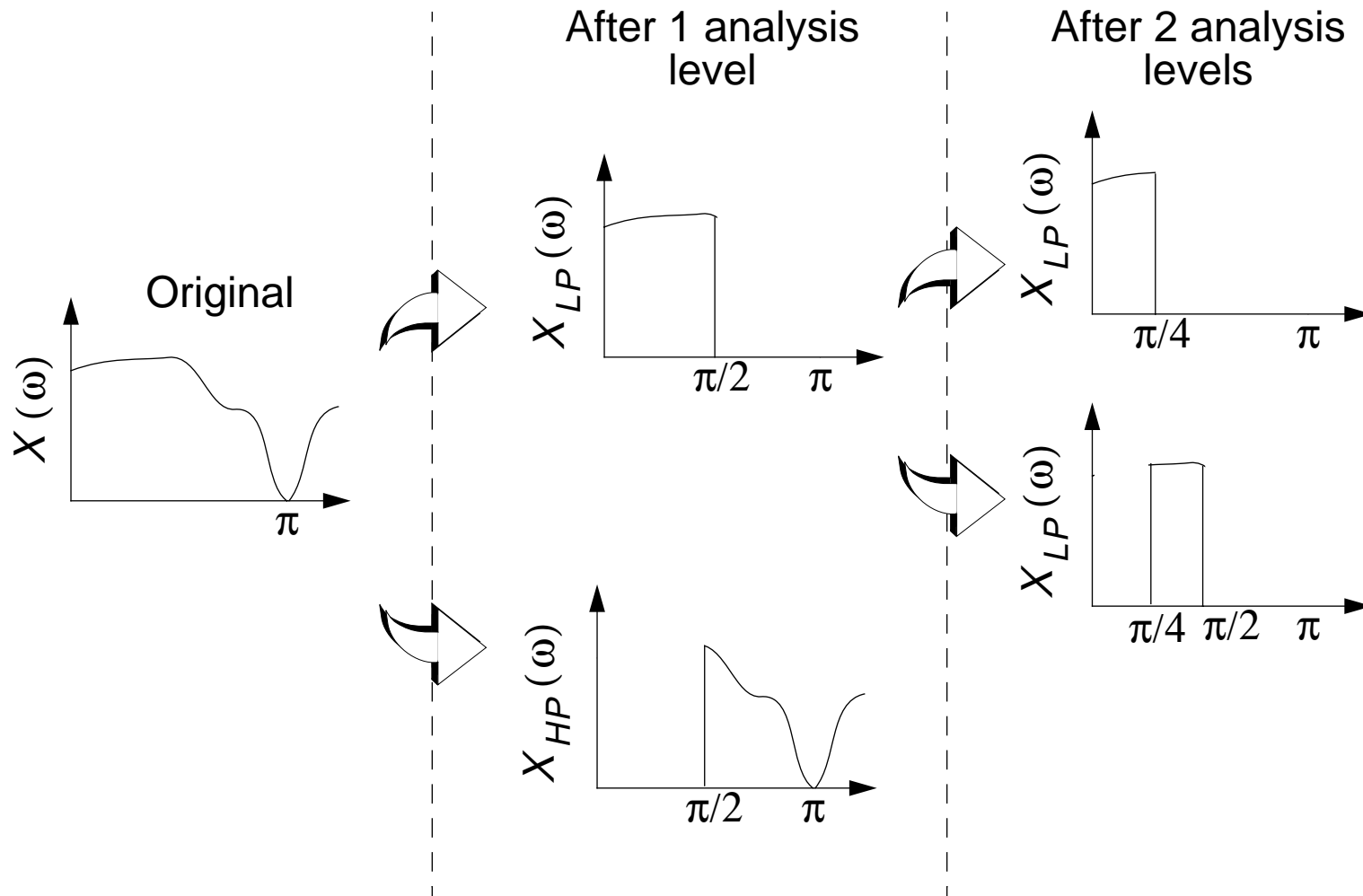
- Most psychophysical experiments measure the VT of individual basis functions, with no or limited spatial masking.
- Exhaustive testing of combinations of luminances and basis functions (“bottom-up test”) is impossible!
- How can we characterize and parameterize spatial masking in natural images? “Top-down test” — use images as the stimuli.

A Wavelet Transform (1 level, 2 band)

scaling (low-frequency) coefficients

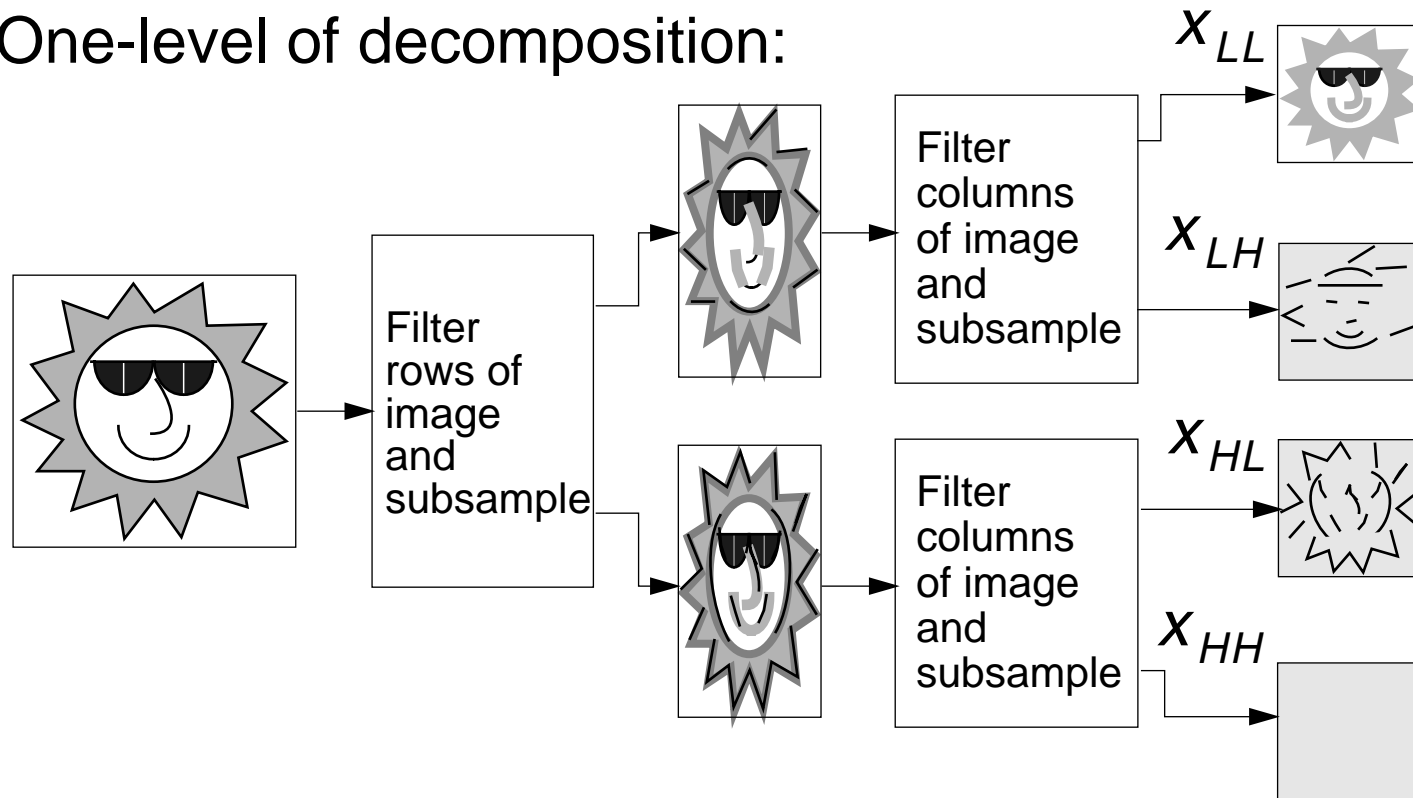


A Frequency-Domain Look

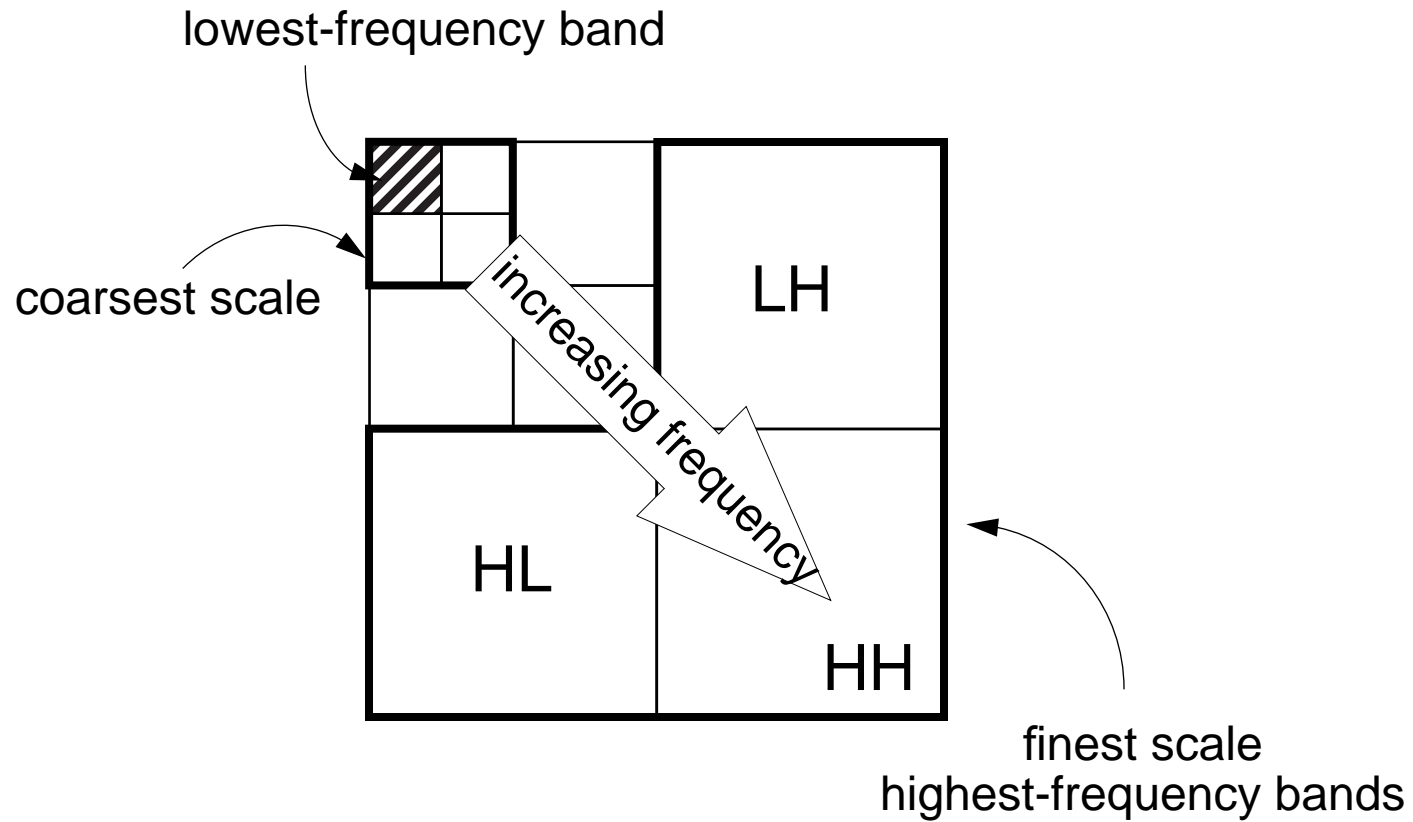


2-D Wavelet Transform

One-level of decomposition:



Nomenclature



Previous Work on *Subthreshold* Wavelet Image Compression

Watson, et. al, Aug. 1997 *IEEE Trans. Image Proc.*

1. Measure visibility threshold of individual wavelet noise basis functions.
2. No spatial masking.
3. No orientation differentiation.
4. Limited number of observers.
5. Resulting quantization matrix provides a single compression ratio.

Quantifying Sensitivity to Supra-threshold Quantization Noise — Our Approach

Measure human sensitivity to uniform quantization noise in wavelet subbands, *within an image*.

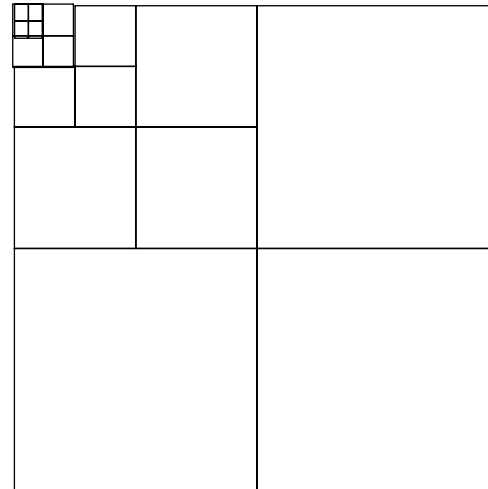
1. Evaluate effects of both *scale & orientation*.
2. Examine visibility thresholds while allowing *spatial masking*.

Definitions

- *Minimum noticeable distortion (MND)* — the distortion at which the VT has been reached.
- *MND step size (MNDSS)* — the quantizer step size at which the MND occurs.
- *1st MNDSS* — the quantizer step size at which there is a MND between the original and the stimulus.
- *Visual distortion unit (VDU)* — unit measuring a minimum-noticeable distortion between two images.

Testing Procedure

- 5-level wavelet transform, 7-9 biorthogonal filters.
- 512×512 images \Rightarrow 1.2 to 19.2 cycles/degree.



- Quantize one band at a time with step sizes

$$q_k = k \times \frac{(\text{dynamic range})}{N}, 0 \leq k \leq N$$

Other bands are not quantized. N is a function of band.

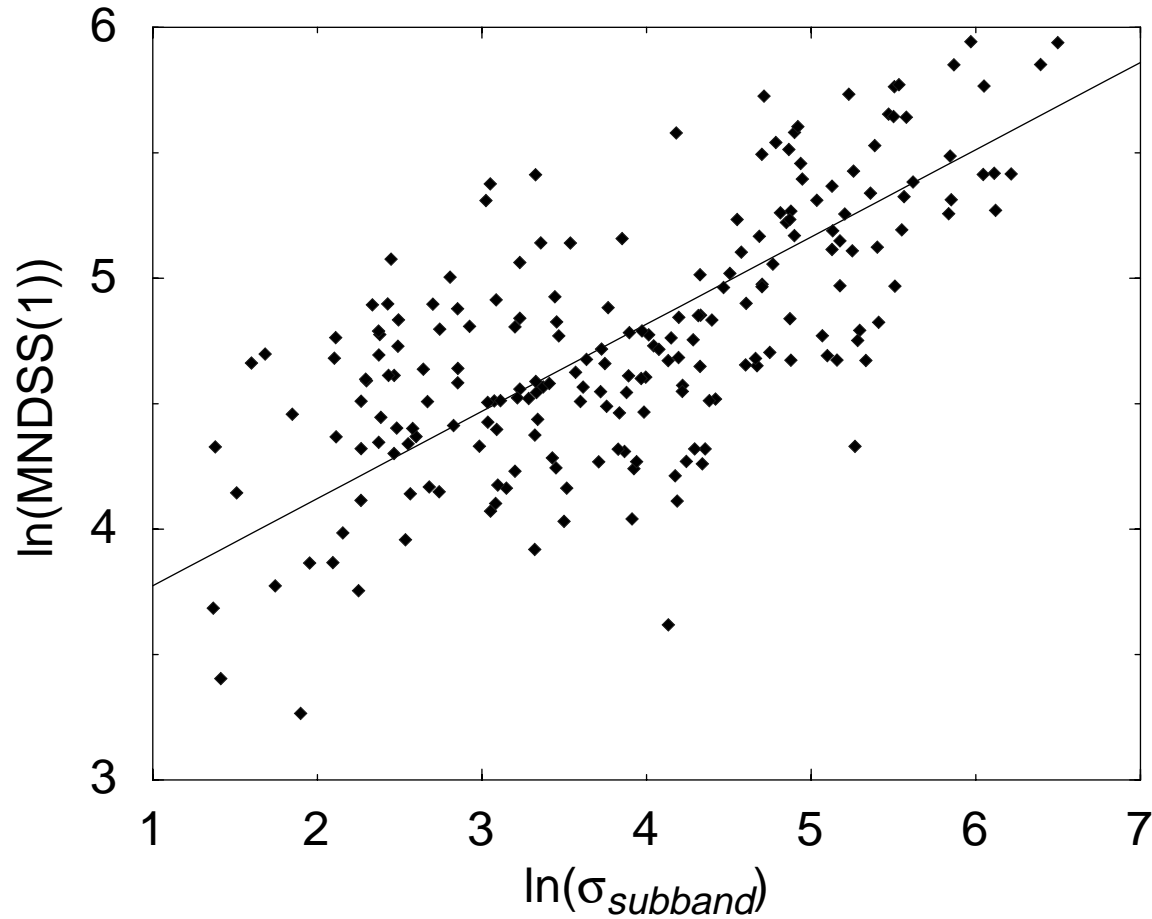
Test Details

- 15 images from a Kodak PhotoCD database, selected based on subject matter, edge/detail content, and AC energy distributions.
- 150 observers (approximately 1 year of testing)
- Each observer goes through 1 test form, evaluating 8 or 9 different subbands from different images (no image is repeated for a single observer to avoid adaptation).

Analysis

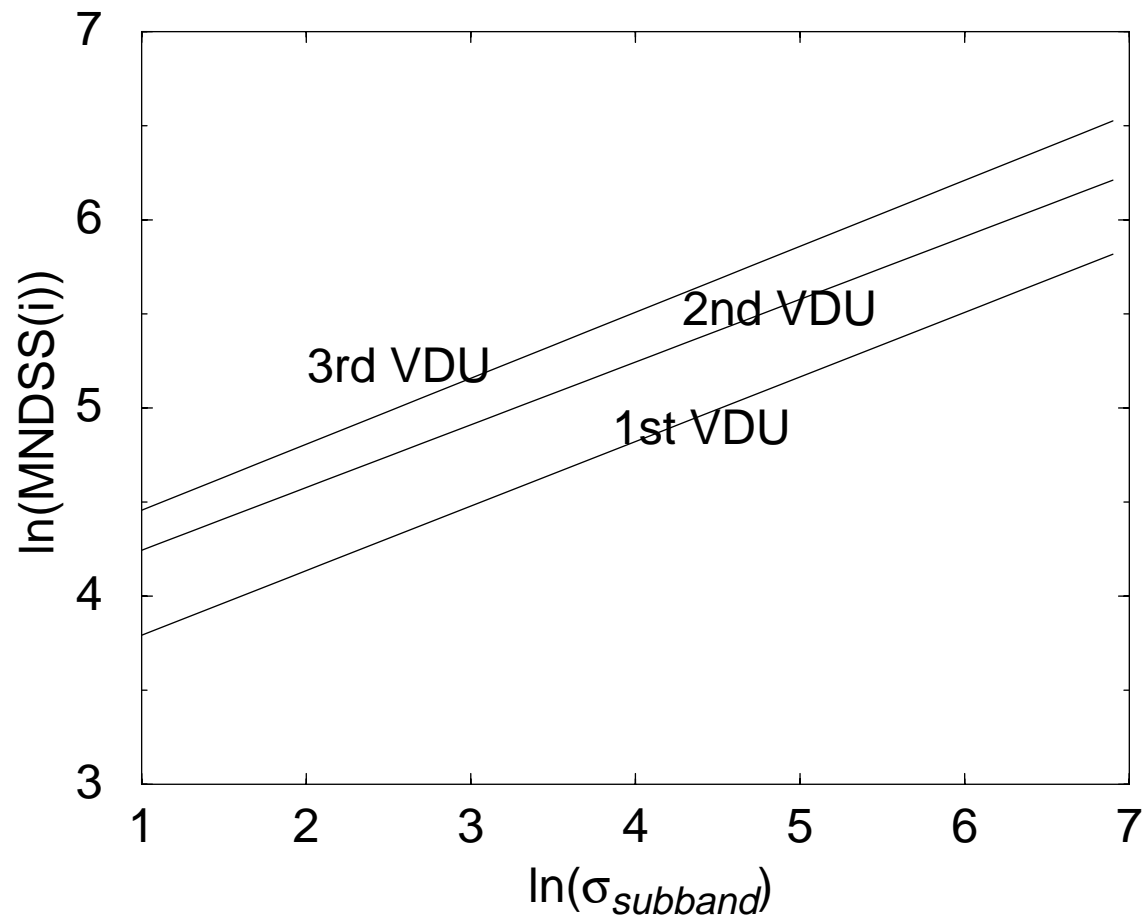
- Parameterize MNDSS as a function of band energy.
- RMS error induced in the image by quantization.
- Contrast sensitivity curves.
- Relative threshold elevations.
- Threshold elevations compared with noise-only stimuli.

MNDSS(1) vs. Band Standard Deviation



Regression fit: $MNDSS = 30.8\sigma^{0.347}$

MNDSS(1-3) vs. Band Standard Deviation



Normalized MND Step Sizes

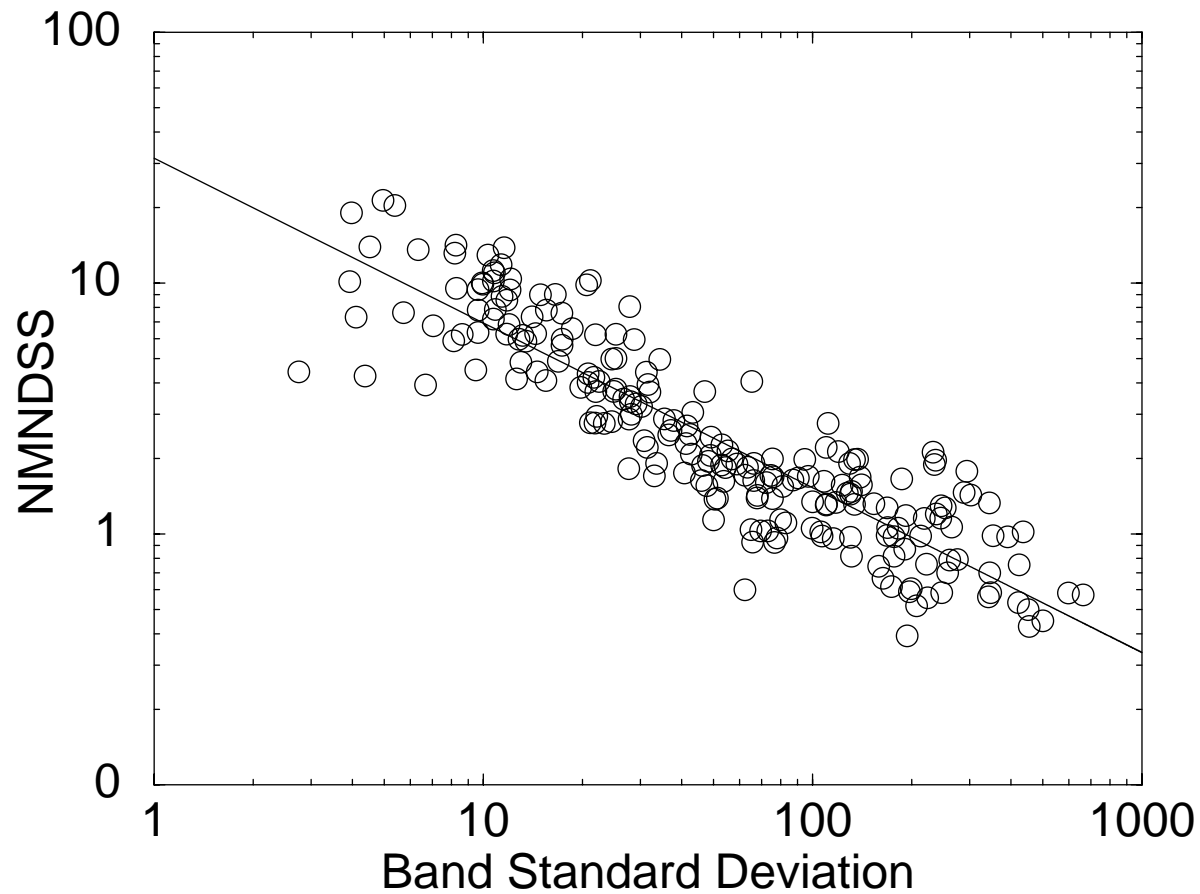
Define the Normalized MND step size as

$$NMNDSS_i = \frac{MNDSS_i}{\sigma_i}$$

This provides a measure of the granularity of the quantizer.

Why? We like to define quantizer step sizes as a multiple of σ (typically $q = 2\gamma\sigma/2^b$).

NMNDSS vs. Band Standard Deviation



Regression fit: $NMNDSS = 30.8\sigma^{-0.653}$

Energy Distribution Classifications (EDCs)

- Subbands in a single scale are classified as having a **horizontal scale EDC (SEDC)** if over 50% of the energy at a scale is in the horizontally oriented (LH) band.

The classification is **vertical SEDC** if over 50% of the energy is in the vertically oriented (HL) band.

Otherwise, the SEDC is called **mixed**.

- An image EDC is defined similarly but considers (high-frequency) subbands at all scales.

SEDC/IEDC Example for *Balloon*

	40%	6.1%	0.6%
35%	10%		
5.3%		1.5%	
0.8%		0.1%	

Finest scale has
SEDC of *vertical*.

Other two scales have
mixed SEDCs.

IEDC is mixed.

Computing the Quantization Error

- Quantizers are NOT operating in the granular region \Rightarrow high-rate/low distortion assumptions do not apply and distortion is NOT $q^2/12$.

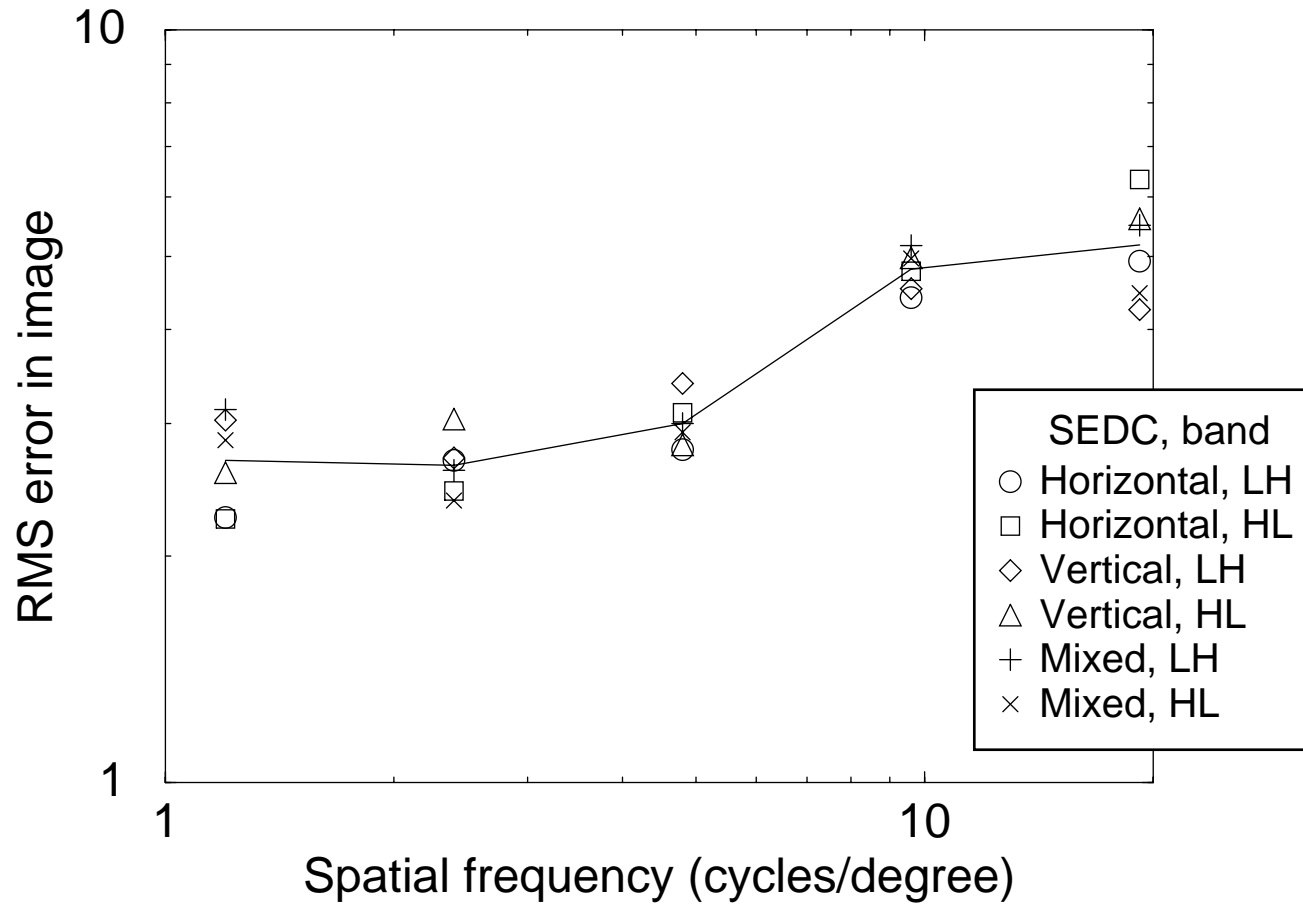
- For $0.5 \leq q/\sigma \leq 5$, can approximate to within 4%

$$MSE \approx \frac{q^2}{12} \left(-0.134 \left(\frac{q}{\sigma} \right) + 1.05 \right)$$

(assume Laplacian distribution for coefficients).

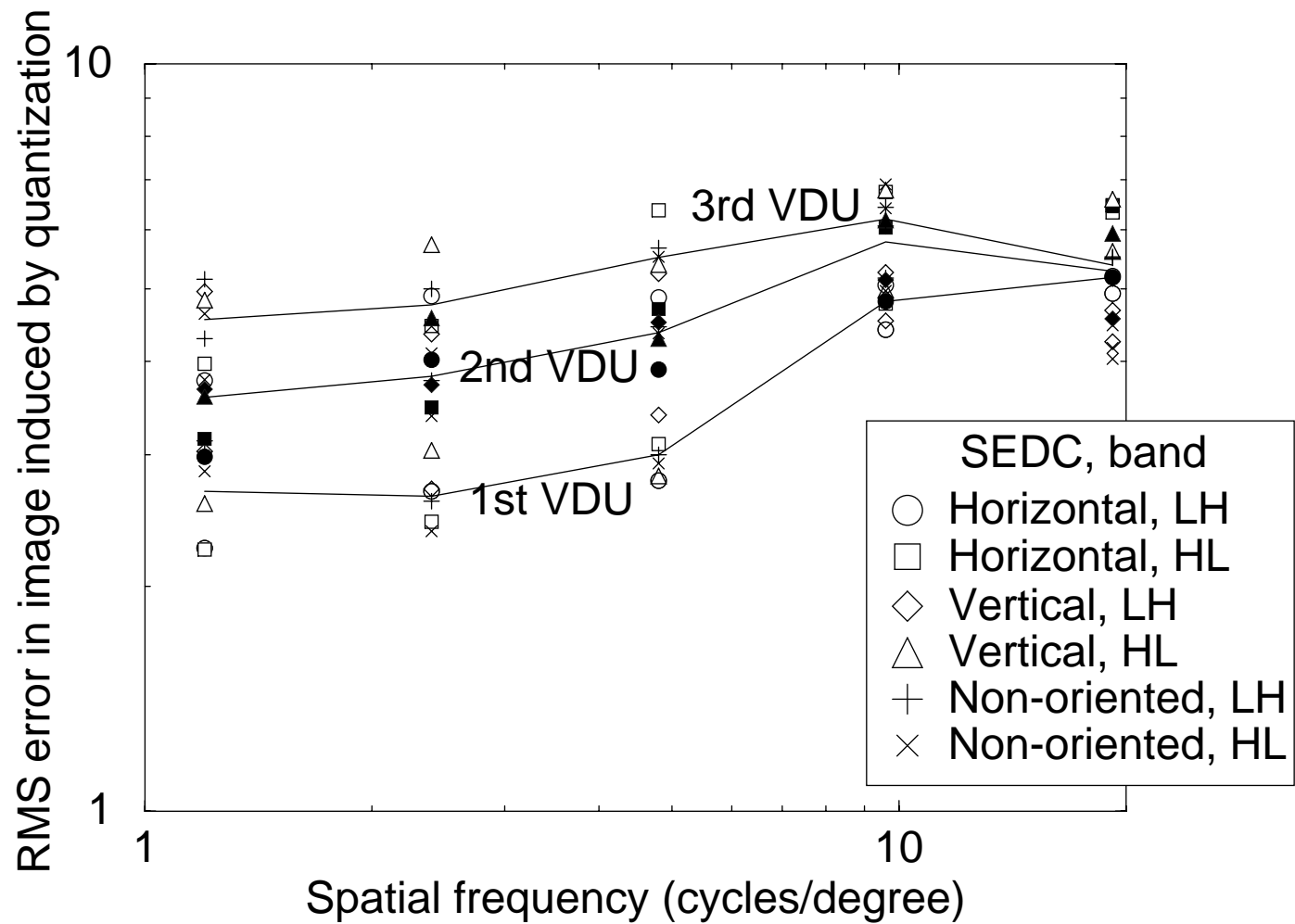
- Bands with different energies and MNDSSs can have the same mean-squared quantization error.

RMS Quantization Error in Image



Constant for a frequency, independent of SEDC.

RMS Quantization Error in Image, VDUs 1-3



Contrast Sensitivity Curves

- The VT occurs when the quantization produces a minimum noticeable distortion in the image.
- Define the contrast threshold measure as

$$CT = \frac{\text{RMS error from quantizing the band}}{\text{RMS energy of the band}}$$

- Intuitively, this represents

$$\frac{\text{average amplitude of distortion}}{\text{average amplitude of signal}}$$

of a frequency in a channel.

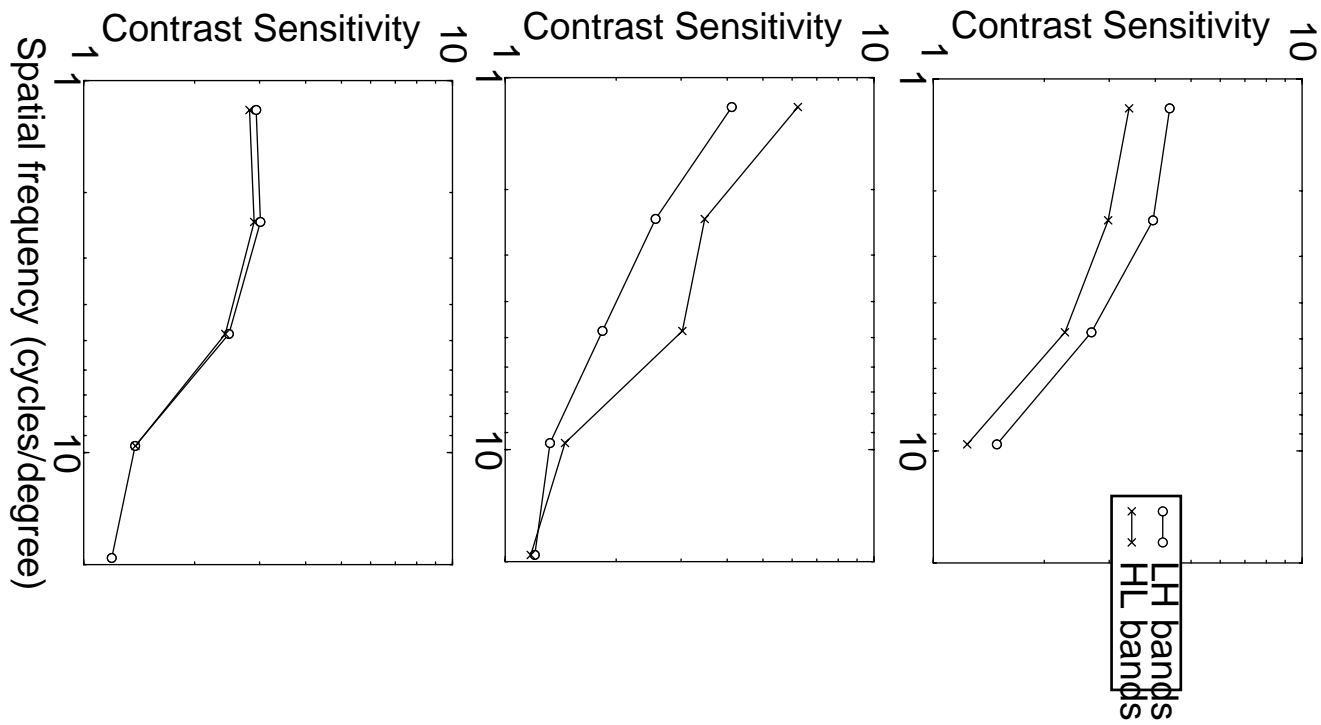
- Produce a CSF for LH & HL bands in each SEDC.

CSF Plots

Mixed

Vertical

Horizontal



Relative Threshold Elevations

- Define the *relative threshold elevation* for a VDU relative to the first VDU as

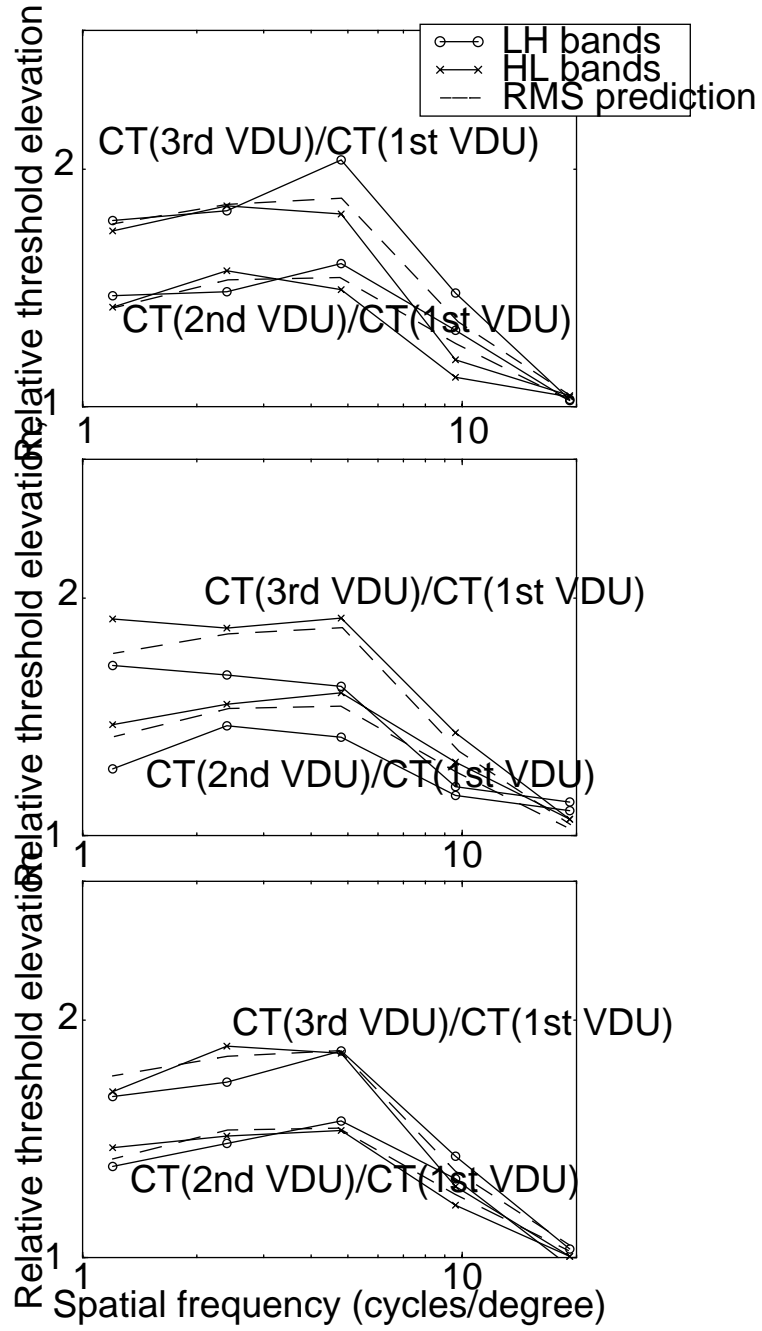
$$RTE (2:1) = \frac{CT (2^{nd} \text{ VDU})}{CT (1^{st} \text{ VDU})}$$

$$RTE (3:1) = \frac{CT (3^{rd} \text{ VDU})}{CT (1^{st} \text{ VDU})}$$

- With RMS error constant for a given frequency and independent of orientation and SEDC, RTE for a given VDU should be equal to the ratio of the RMS curves.
- This is observed!

RTE Plots

Non-oriented Vertical Horizontal



Threshold Comparisons

- Compare with noise-only stimuli in single subbands (Watson).

- For spatial frequencies of 2-10 cycles/degree,

$$VT_{\text{complex stimuli}} \approx 2.5 VT_{\text{noise-only stimuli}}$$

- For spatial frequencies around 20 cycles/degree, the thresholds are approximately the same.

$$VT_{\text{complex stimuli}} \approx VT_{\text{noise-only stimuli}}$$

- Spatial masking at work!

Analysis Summary

- MNDSS(1-3) strongly influenced by the subband energy.
- Equal visual distortion implies equal RMS energy at a given frequency; overload operation of quantizers implies that the MNDSS is a function of subband energy.
- CSFs demonstrate orientation-specific spatial masking in images.
- Constant RMS (MSE) predicts relative threshold elevations well.

A Quantization Strategy

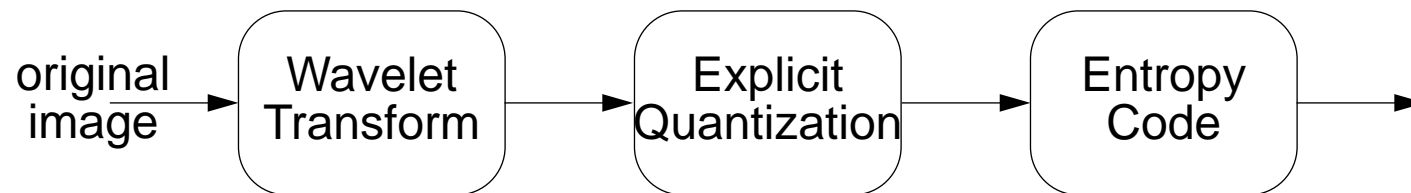
- Assumptions:
 - Quantization with a step size of $\alpha \times MNDSS$ produces α VDU.
 - Additive distortion.
- To achieve an image with which differs from the original by 1 VDU, quantize each band i with

$$Q_i = \frac{K}{\sigma_i} MNDSS(\sigma_i) = Ka\sigma_i^{b-1}$$

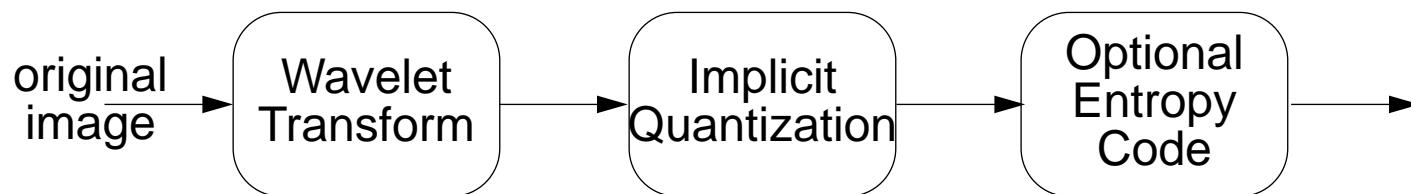
where a and b from the regression line and K is selected so $\sum_i K/\sigma_i = 1$.

Incorporation into Wavelet-Based Image Compression

1. Comparison with prior work on subthreshold perception using a basic wavelet coder.



2. Application to embedded coders (e.g. SPIHT).



Important Points

- We to react to equal MSE as a function of frequency but not orientation.
- Scaling quantization step sizes to achieve a given rate is supported by experimental data. But you have to scale the right step sizes!
- Using scaled MNDSS(1)'s produces images with visible differences compared with using scaled sub-threshold step sizes.

Subjective Quality Evaluation of Low Bitrate Video

- 8 sequences, each 30 seconds long (2 high, 4 medium, 2 low motion)
- 3 motion-compensated coders
 - Sorensen (Quicktime)
 - H.263+
 - Wavelet-based RD-optimized
- 5 bitrate/frame rate combinations
 - 1 “high” quality, 1 “low” quality
 - 3 at same bitrate but different frame rates

Psychophysical Experiment

- 19 observers watched all (8 sequences) x (3 coders) x (5 bitrate/frame rate combos) = 120 sequences, presented randomly.
- Single stimulus continuous quality evaluation (SSCQE): observers continuously evaluated quality using a linear slider.
- Raw data is 15 continuous waveforms for each sequence.

Varying Frame Rate at Fixed Bitrate

- Perceived quality **INCREASES** with **DECREASING** frame rate for fast, jerky-motioned sequences (*car chase, martial arts*).
- Perceived quality is **CONSTANT** with **DECREASING** frame rate for smooth, full-frame motion, or no motion.