Wavelet Denoising and Speech Enhancement

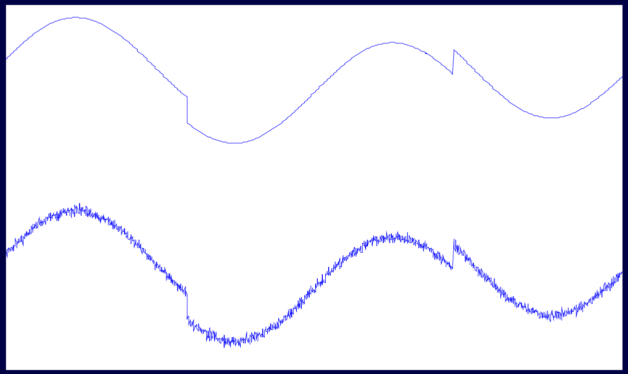
Balakrishnan V

Nash Borges

Luke Parchment

Motivation

- Signals and Images are often corrupted by noise
  - thermal channel noise, background noise, etc
- Follow on algorithms (classifiers, compression) work best on clean data



Thresholding

- Signals of interest (images, speech, etc) have redundancy which results in sparse wavelet coefficients
- AWGN has no such redundancy and therefore is not compressible
  - Noisy samples result in noisy wavelet coefficients
- Properly thresholding the wavelet coefficients of a signal corrupted by noise can remove the noise and preserve the signal

Noise Estimation

- Most denoising algorithms require an estimate of the noise variance
- For images, we can estimate the noise level from the highpass coefficients
- If you assume the noise is Gaussian, the robust median estimate applies asymptotically:

$$\sigma = \frac{\text{median}\left(\left|HH_1\right|\right)}{0.6745}$$

Thresholding Algorithms

- VisuShrink
  - D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?" Journal of the Royal Statistics Society, Series B, vol. 57, pp. 301-369, 1995.
- SureShrink
  - D. L. Donoho, and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage." J. Am. Statist. Ass., vol. 90, no. 432, pp 1200-1224, 1995.
- BayesShrink
  - G. Chang, B. Yu and M. Vetterli "Adaptive Wavelet Thresholding for Image Denoising and Compression." IEEE Trans of Image Processing, vol. 9, no. 9, pp1532-1546, 2000.

VisuShrink

- Minimizes the probability that any noise sample will exceed the threshold
- Given a sequence of Gaussian random variables, the expected value of the maximum increases with the length of the signal

$$\lambda_{\mathrm{UNIV}} = \sigma\sqrt{2\log N}$$

- Same threshold applied to all decomposition levels

SureShrink

- Stein's unbiased risk estimate (SURE)
- For each subband, find the threshold:

$$\lambda_{SURE} = \arg\min_{0 \le \lambda \le \sqrt{\log d}} \mathrm{SURE}(\lambda; x)$$

which minimizes the risk:

$$\mathrm{SURE}(\lambda; x) = d - 2\#\left(i : |x_i| \le \lambda\right) + \sum_{i=1}^{d} \min\left(|x_i|, \lambda\right)^2$$

BayesShrink

- Assuming generalized Gaussians, for <u>each subband</u> find the threshold

$$\lambda_{\text{BAYES}} = \frac{\hat{\sigma}^2_{\text{noise}}}{\hat{\sigma}_{\text{signal}}}$$

that minimizes the Bayesian risk

$$r(\lambda) = E\left(\hat{X} - X\right)^2 = E_X E_{Y|X}\left(\hat{X} - X\right)^2$$

- <u>Nearly</u> optimal and independent of shaping parameter β

Does not threshold the LL_n coefficients

BayesShrink

- Assuming no signal in HH_1, estimate the noise variance from these coefficients
  - Given the model and assuming GGD,

    $$Y = \text{signal} + \text{noise} \quad \Rightarrow \sigma_Y^2 = \sigma_{signal}^2 + \sigma_{noise}^2$$

    for each subband

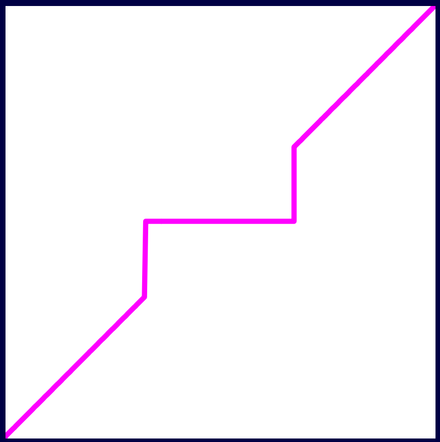
$$\hat{\sigma}_{signal} = \sqrt{\max\left(\hat{\sigma}_Y^2 - \hat{\sigma}_{noise}^2, 0\right)}$$

$$\hat{\sigma}_Y^2 = \frac{1}{d}\sum_{i=1}^{d} Y_i^2 \qquad \sigma_{noise}^2 = \left(\frac{\text{median}\left(\left|HH_1\right|\right)}{0.6745}\right)^2$$

Thresholding

Hard Thresholding

$$W(Y,\lambda) = Y \cdot u\left(\left|Y\right| - \lambda\right)$$



Soft Thresholding

$$W(Y,\lambda) = \operatorname{sign}(Y)\max\left(0,\left|Y\right| - \lambda\right)$$

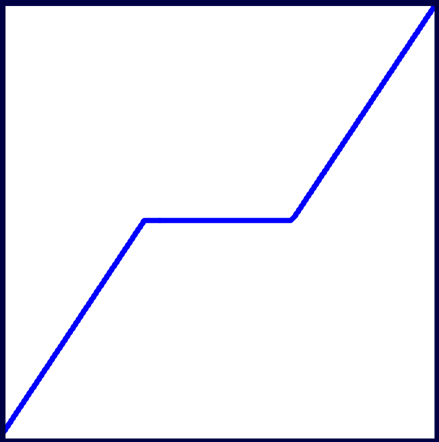
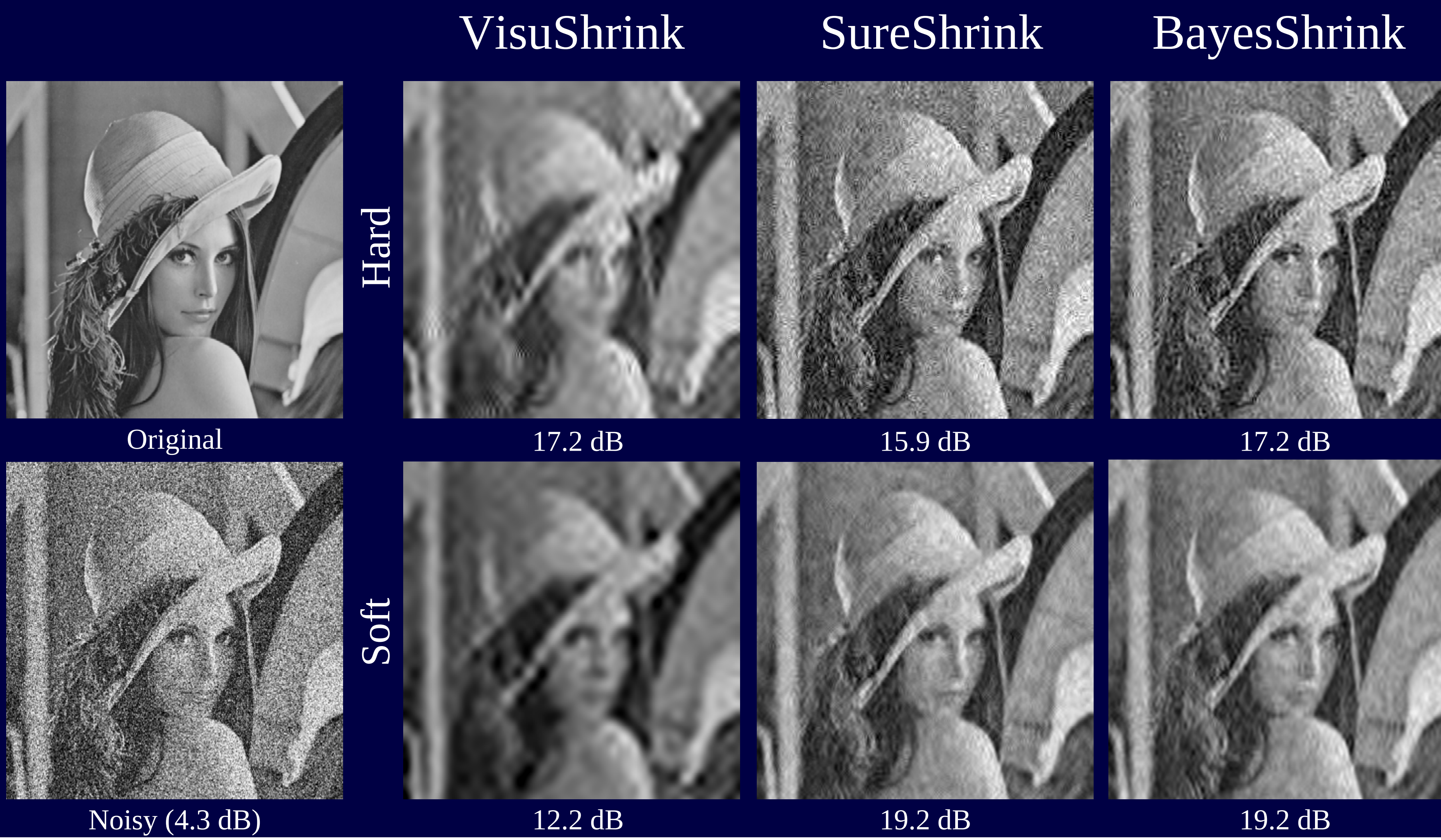


Image Denoising Example



1D Results

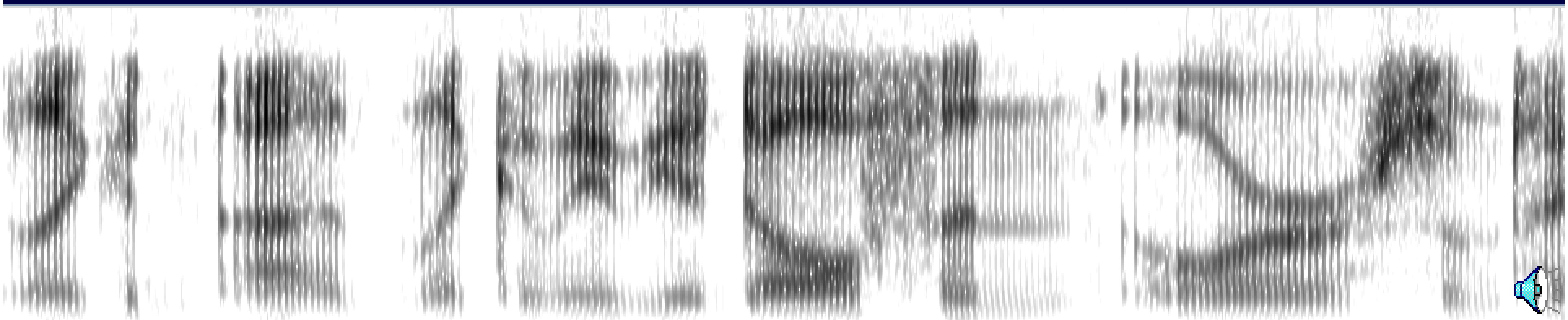
SNRs		VisuShrink		SureShrink		BayesShrink	
	Noisy Signal	Hard	Soft	Hard	Soft	Hard	Soft
blocks							
var = 0.001	23.4	28.8	21.2	25.7	**29.0**	24.1	27.5
var = 0.01	13.4	19.0	12.9	16.9	**20.4**	17.8	20.1
var = 0.1	3.4	10.3	5.6	10.2	**13.1**	11.7	**13.1**
bumps							
var = 0.001	13.0	18.7	13.7	16.4	**20.0**	16.3	18.7
var = 0.01	3.0	10.7	6.9	8.6	**12.1**	9.2	11.2
var = 0.1	-7.0	3.2	1.5	1.2	**5.3**	2.1	3.4
heavisine							
var = 0.001	24.1	32.8	23.5	30.0	**33.8**	31.0	33.1
var = 0.01	14.1	23.2	14.1	21.7	24.8	24.1	**24.9**
var = 0.1	4.1	10.9	5.3	12.3	15.5	15.1	**15.7**
doppler							
var = 0.001	25.3	32.3	23.8	29.7	**32.6**	28.5	30.8
var = 0.01	15.3	22.9	14.7	20.7	**23.8**	20.8	23.0
var = 0.1	5.3	12.3	6.0	12.3	**15.2**	13.4	**15.2**

2D Results

SNRs		VisuShrink		SureShrink		BayesShrink	
	Noisy Signal	Hard	Soft	Hard	Soft	Hard	Soft
lena							
var = 0.001	24.3	25.6	22.7	26.6	**28.9**	26.8	28.8
var = 0.01	14.3	20.9	18.1	20.7	**23.7**	21.4	23.6
var = 0.1	4.3	17.2	12.2	15.9	**19.2**	17.2	**19.2**
boat							
var = 0.001	24.7	23.5	20.9	26.2	**27.6**	26.4	27.5
var = 0.01	14.7	19.2	17.0	19.3	**22.2**	19.8	22.1
var = 0.1	4.7	16.2	12.1	14.5	**17.8**	15.8	**17.8**
tank							
var = 0.001	24.2	22.3	20.6	25.3	**26.6**	25.4	**26.6**
var = 0.01	14.2	19.6	17.8	19.4	**21.9**	20.0	**21.9**
var = 0.1	4.2	18.2	12.4	15.8	18.8	17.2	**18.9**
aerial							
var = 0.001	27.2	23.8	21.0	27.9	**29.2**	28.0	**29.2**
var = 0.01	17.2	19.1	17.2	20.1	**22.8**	20.7	22.7
var = 0.1	7.2	16.2	13.3	15.1	**18.2**	16.0	**18.2**

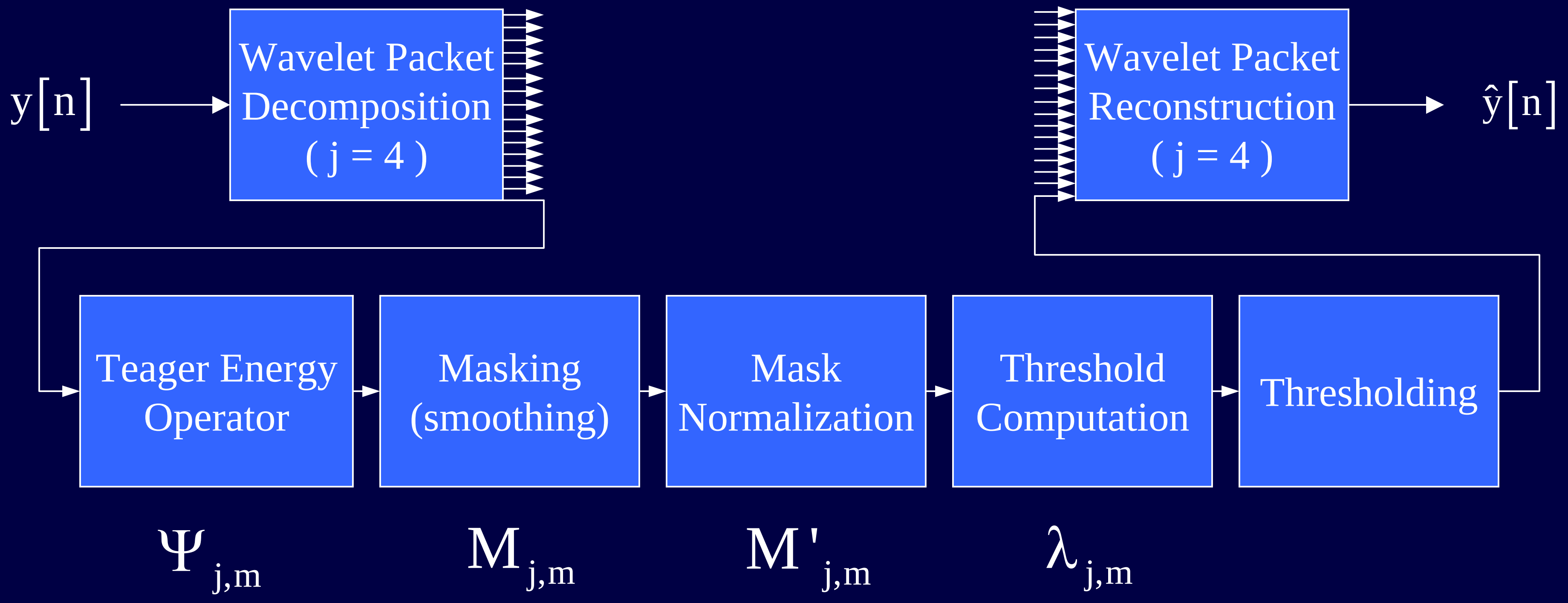
Speech Enhancement

- Unlike images, band-limited speech occupies most of the frequency range



- Want higher thresholds for noise/silence
- Lower/zero threshold for speech
  - preserve speech quality/intelligibility

Time-Adapted Thresholding



Masking Speech

- Teager Energy Operator

$$\Psi_{j,m} = w_{j,m}^2 - w_{j,m-1} \cdot w_{j,m+1}$$

- Smooth TEOs with window

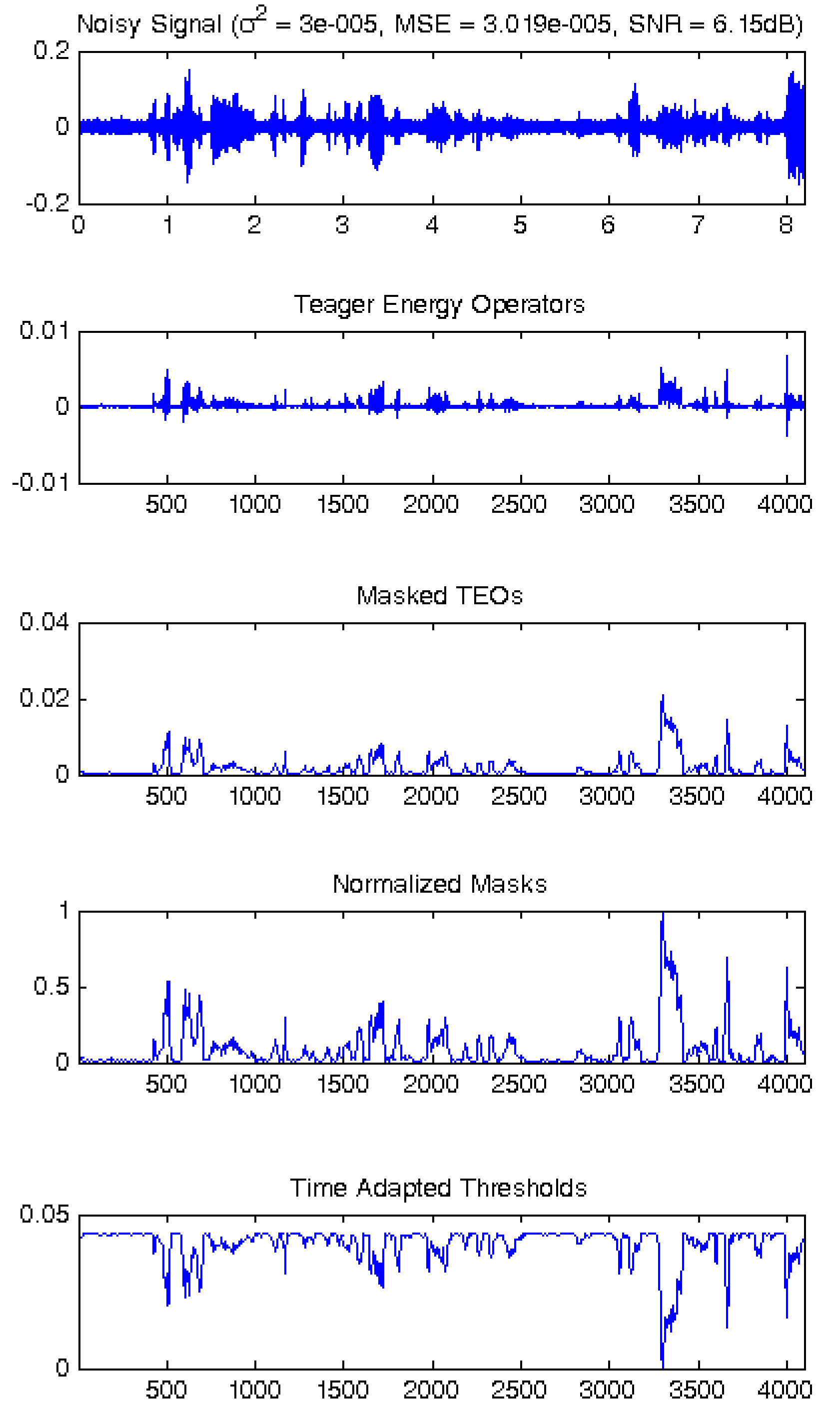
$$M_{j,m} = \Psi_{j,m} * h_j$$

- Normalize smoothed TEOs

$$M'_{j,m} = \frac{M_{j,m}}{\max\left(M_{j,m}\right)}$$

- Time-Adapted Thresholds

$$\lambda_{j,m} = \lambda_j\left(1 - M'_{j,m}\right)$$



Time Adapted Thresholding

- Level-dependent noise estimate

$$\hat{\sigma}_j = \frac{\text{median}\left(\left|Y_j - \text{median}\left(Y_j\right)\right|\right)}{0.6745}$$

- Level-dependent universal threshold

$$\lambda_j = \hat{\sigma}_j \sqrt{2 \log d}$$

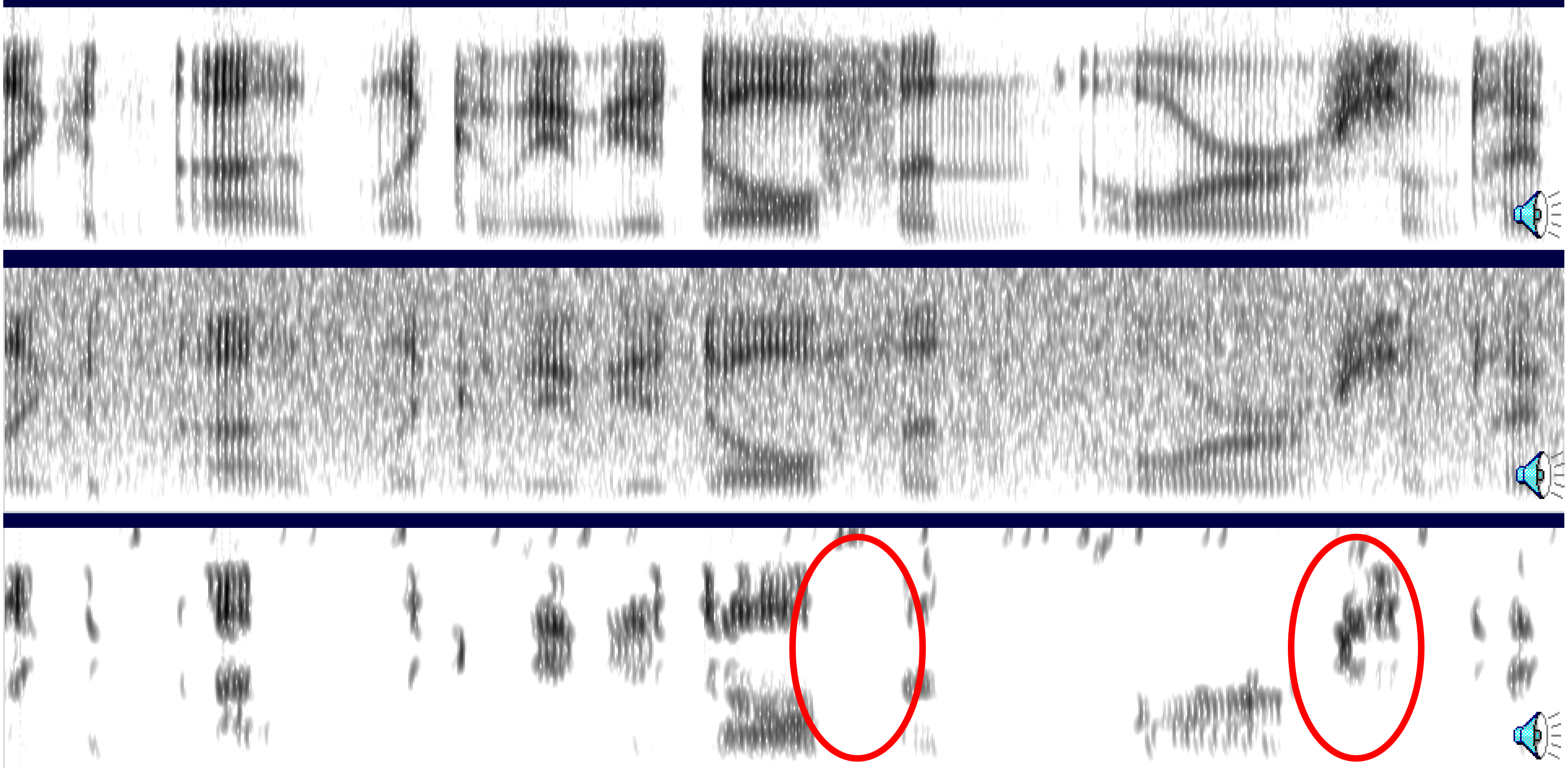
- Time-adapted threshold

$$\lambda_{j,m} = \lambda_j \left(1 - M'_{j,m}\right)$$

- Soft thresholding, point by point

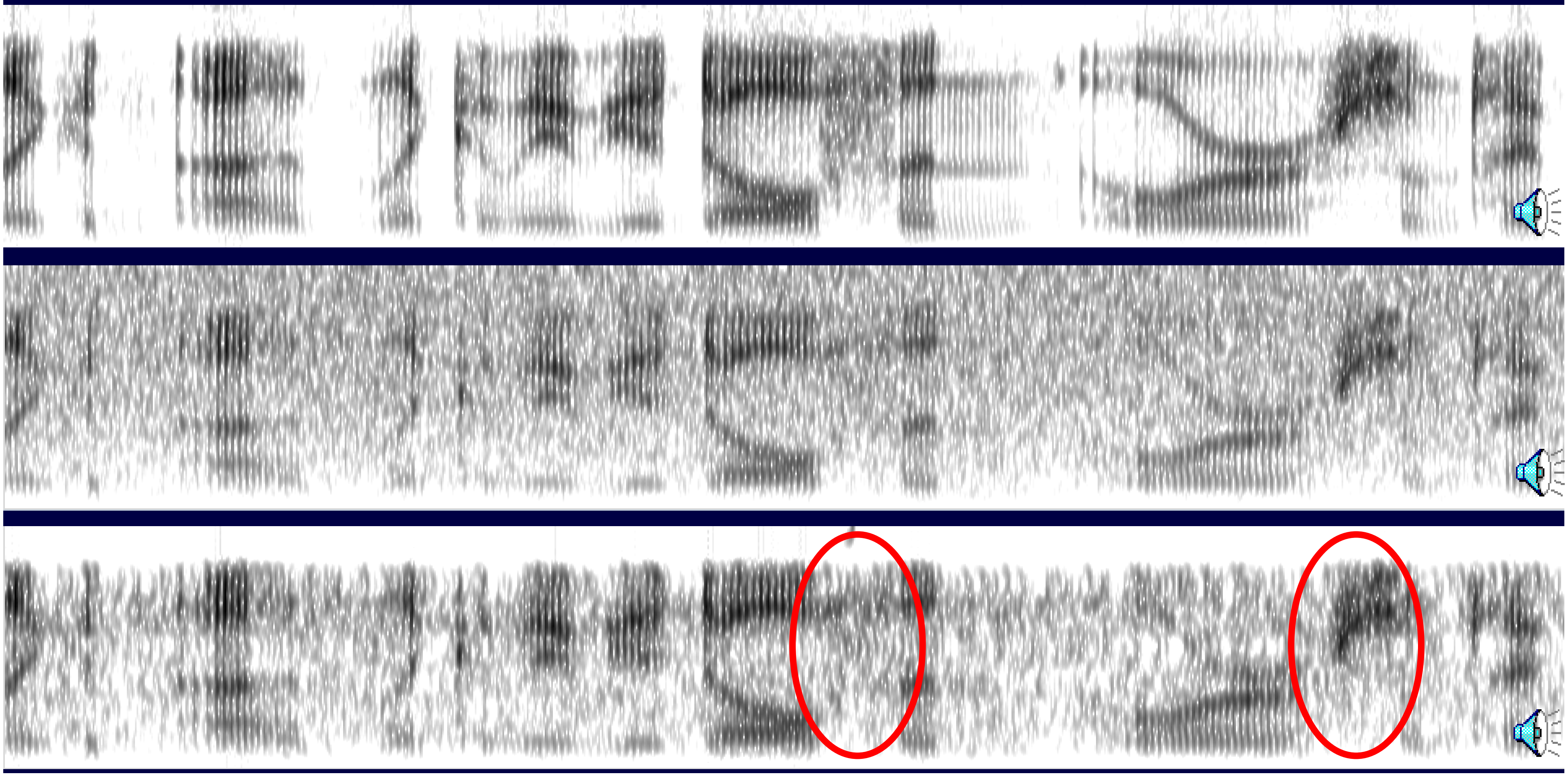
Distortion Problems

- Universal-based thresholding can be overly aggressive leading to distortions

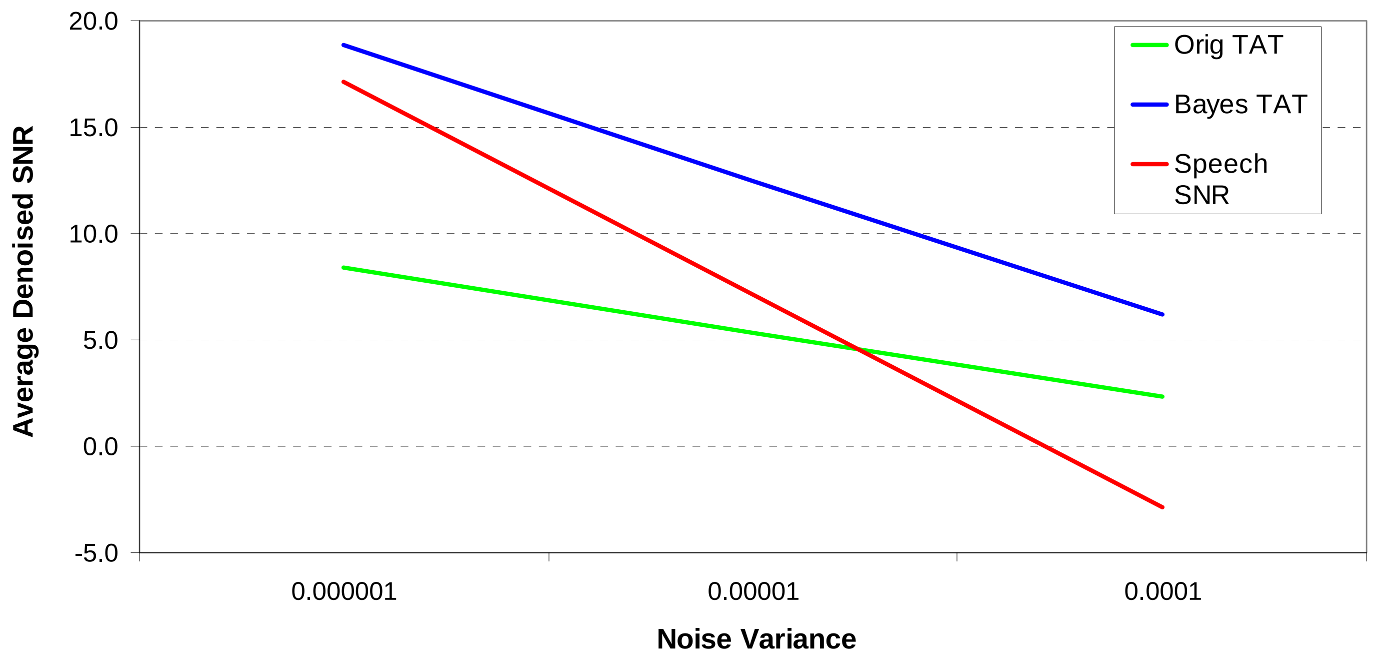


What about TAT w/ BayesShrink?

- BayesShrink leaves slightly more noise, but preserves speech intelligibility



Speech Enhancement Results



Speech Enhancement References

– M. Bahoura and J. Rouat, "Wavelet Speech Enhancement Based on the Teager Energy Operator." IEEE Sig. Proc. Letters, vol. 8, no. 1, pp. 10-12, 2001.
– S. Chen and J. Wang, "Speech Enhancement Using Perceptual Wavelet Packet Decomposition and Teager Energy Operator." Journal of VLSI Sig. Proc., vol. 32, pp. 125-139, 2004.

Demo

Avoiding Edge Effects

- Mirror the image/signal before applying DWT since we don’t care about compression
- Throw away extra after IDWT