Speech Modeling and Enhancement in Nonstationary Noise Environments Part II

Prof. Israel Cohen and Prof. Sharon Gannot

Elect. Eng. Dept., Technion - Israel Inst. of Tech., Israel
School of Engineering, Bar-Ilan University, Israel

SIPA 2011

Outline

1. **Noise Estimation**
   - Minima Controlled Recursive Averaging (MCRA)
   - Minimum Statistics (MS)
   - Implementation
   - Experimental Results
2. **Statistical Models**
   - Spectral Analysis
   - GARCH Model
   - Speech Modeling
3. **Variance Estimation**
   - Relation to Decision-Directed Estimation
   - Experimental Results
4. **Conclusions**

Reminder: Problem Formulation

Let $\{Y_{tk}\}$ denote a noisy speech signal in the STFT domain:

$$\begin{aligned} H_1^{tk}\,(\text{speech present}): \quad Y_{tk} &= X_{tk} + D_{tk} \\ H_0^{tk}\,(\text{speech absent}): \quad Y_{tk} &= D_{tk}\,. \end{aligned}$$

The spectral enhancement problem can be formulated as

$$\min_{\hat{X}_{tk}} E\left\{ d\left(X_{tk}, \hat{X}_{tk}\right) \,\middle|\, \hat{p}_{tk}\,, \hat{\lambda}_{tk}\,, \widehat{\sigma^2_{tk}}\,,\ Y_{tk} \right\}$$

- $d\left(X_{tk}, \hat{X}_{tk}\right)$ - distortion measure between X_{tk} and $\hat{X}_{tk}$
- $\hat{p}_{tk} = P\left(H_1^{tk} \mid \psi_t\right)$ - speech presence probability estimate
- $\hat{\lambda}_{tk} = E\left\{|X_{tk}|^2 \mid H_1^{tk},\, \psi_t\right\}$ - speech spectral variance estimate
- $\widehat{\sigma^2_{tk}} = E\left\{|Y_{tk}|^2 \mid H_0^{tk},\, \psi_t\right\}$ - noise spectral variance estimate
- ψ_t - information employed for estimation at frame t (*e.g.*, noisy data observed through time t)

Noise Spectrum Estimation

Minima Controlled Recursive Averaging (MCRA)

- A common noise estimation technique is to recursively average past spectral power values of the noisy measurement during periods of speech absence:

$$H_0^{tk} : \bar{\sigma}^2_{t+1,k} = \alpha_d \, \bar{\sigma}^2_{tk} + (1 - \alpha_d) \, |Y_{tk}|^2$$
$$H_1^{tk} : \bar{\sigma}^2_{t+1,k} = \bar{\sigma}^2_{tk}$$

  where α_d $(0 < \alpha_d < 1)$ denotes a smoothing parameter.
- Under speech presence uncertainty

$$\begin{aligned}\bar{\sigma}^2_{t+1,k} &= \tilde{p}_{tk} \, \bar{\sigma}^2_{tk} \\ &\quad + (1 - \tilde{p}_{tk}) \left[\alpha_d \, \bar{\sigma}^2_{tk} + (1 - \alpha_d) \, |Y_{tk}|^2\right]\end{aligned}$$

  where $\tilde{p}_{tk}$ is an estimator for the conditional speech presence probability $p_{tk} = P\left(H_1^{tk} \mid Y_{tk}\right)$.

- Equivalently

$$\bar{\sigma}^2_{t+1,k} = \tilde{\alpha}_{tk}\,\bar{\sigma}^2_{tk} + (1 - \tilde{\alpha}_{tk})\,|Y_{tk}|^2$$

where

$$\tilde{\alpha}_{tk} \triangleq \alpha_d + (1 - \alpha_d)\,\tilde{p}_{tk}$$

is a time-varying frequency-dependent smoothing parameter, adjusted by the speech presence probability.

- Deciding speech is absent (H_0) when speech is present (H_1) is more destructive when estimating the speech than when estimating the noise.
- Hence, we make a distinction between the estimator $\hat{p}_{tk}$ used for estimating the clean speech, and the estimator $\tilde{p}_{tk}$, which controls the adaptation of the noise spectrum. Generally $\hat{p}_{tk} \geq \tilde{p}_{tk}$.

- The estimator $\tilde{p}_{tk|t}$ is biased toward higher values, since deciding speech is absent when speech is present results ultimately in the attenuation of speech components.
- Accordingly, we include a bias compensation factor in the noise estimator

$$\hat{\sigma}^2_{t+1,k} = \beta \cdot \bar{\sigma}^2_{t+1,k}$$

such that the factor β ($\beta \geq 1$) compensates the bias when speech is absent:

$$\beta \triangleq \left. \frac{\sigma^2_{tk}}{E\left\{\bar{\sigma}^2_{tk}\right\}} \right|_{H_0} .$$

- The value of β is completely determined by the particular estimator for the *a priori* speech absence probability.

Minimum Statistics

- Let α_s $(0 < \alpha_s < 1)$ be a smoothing parameter, and let b denote a normalized window function of length $2w+1$, *i.e.*, $\sum_{i=-w}^{w} b_i = 1$.
- The frequency smoothing of the noisy power spectrum in each frame is defined by

$$S_{tk}^{f} = \sum_{i=-w}^{w} b_i \left|Y_{t,k-i}\right|^2 .$$

- Subsequently, smoothing in time is performed by a first order recursive averaging:

$$S_{tk} = \alpha_s S_{t-1,k} + (1-\alpha_s) S_{tk}^{f} .$$

- The minima values of S_{tk} are picked within a finite window of length D, for each frequency bin:

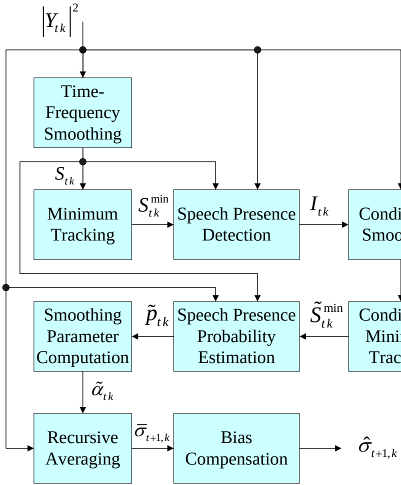
$$S_{tk}^{\min} \triangleq \min\left\{S_{t',k} \mid t-D+1 \leq t' \leq t\right\} .$$

- It follows that there exists a constant factor $B_{\min}$, independent of the noise power spectrum, such that

$$E\left\{S_{tk}^{\min} \mid H_0\right\} = B_{\min}^{-1} \cdot \sigma_{tk}^2 .$$

- The factor $B_{\min}$ represents the bias of a minimum noise estimate, and generally depends on the values of D, α_s, b and the spectral analysis parameters (type, length and overlap of the analysis windows)
- The value of $B_{\min}$ can be estimated by generating a white Gaussian noise, and computing the inverse of the mean of $S_{tk}^{\min}$.

Block diagram of the IMCRA noise estimator



Implementation

A free MATLAB code is available on:
http://www.ee.technion.ac.il/people/IsraelCohen/

Initialization at the first frame for all frequency-bins $k = 1, \ldots, N/2$:

$$\hat{\sigma}^2_{0k} = |Y_{0k}|^2; \quad \bar{\sigma}^2_{0k} = |Y_{0k}|^2; \quad S_{0k} = S^f_{0k}; \quad S^{\min}_{0k} = S^f_{0k};$$

For all short-time frames $t = 0, 1, \ldots$

For all frequency-bins $k = 1, \ldots, N/2$

1) Compute the *a posteriori* SNR $\hat{\gamma}_{tk}$ and the *a priori* SNR $\hat{\xi}_{tk}$ with the initial condition $\hat{\xi}_{0k} = \alpha + (1-\alpha)\max\{\hat{\gamma}_{0k} - 1, 0\}$.

2) Compute the conditional spectral estimate under the hypothesis of speech presence $\hat{X}_{tk|H_1} = G_{\mathrm{LSA}}(\hat{\xi}_{tk}, \hat{\gamma}_{tk})\, Y_{tk}$.

3) Compute the smoothed power spectrum S_{tk} and update its running minimum: $S_{tk}^{\min} = \min \left\{ S_{t-1,k}^{\min} ,\ S_{tk} \right\}$.

4) Compute the speech presence probability $\tilde{p}_{tk}$, and the smoothing parameter $\tilde{\alpha}_{tk}$.

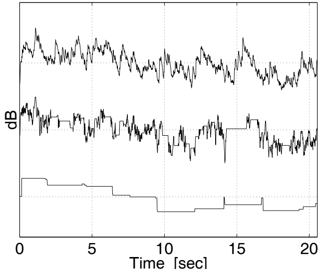
5) Update the noise spectrum estimate $\hat{\sigma}_{t+1,k}^2$.

6) Compute the speech presence probability $\hat{p}_{tk}$.

7) Compute the speech spectral estimate $\hat{X}_{tk}$.

Example 1

Noise estimation, F16 cockpit noise at 0 dB segmental SNR, a single frequency bin $k = 40$ (center frequency 1219 Hz):

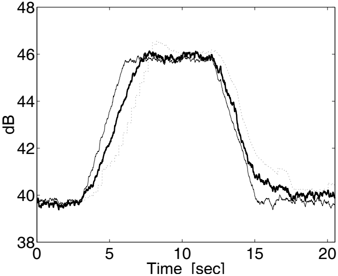


Ideal (top), IMCRA (center), and MS (bottom) noise estimates (top and bottom graphs are displaced by ± 10 dB, for clarity).

Clearly, the IMCRA noise estimate follows the noise power more closely than the MS noise estimate.

Example 2

Non-stationary WGN at 0 dB segmental SNR:

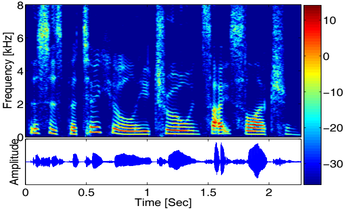


Ideal (fine), IMCRA (heavy), and MS (dotted) average noise estimates.

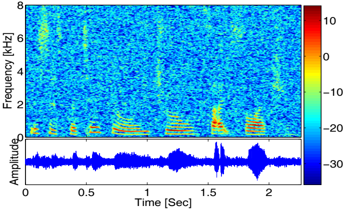
The response of the IMCRA estimator to increasing or decreasing noise power is essentially much faster than that of the MS estimator, due to the recursive averaging mechanism.

Results

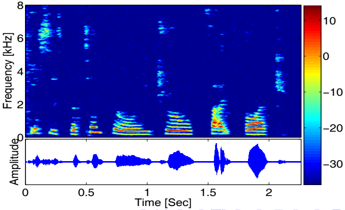
Clean speech signal



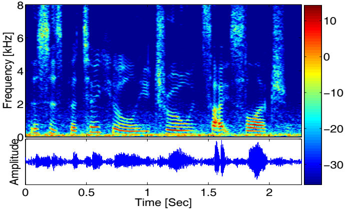
Noisy signal, white Gaussian noise
LSD = 12.67dB, PESQ= 1.74



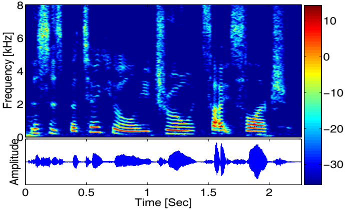
Enhanced signal, OM-LSA
LSD = 5.10dB, PESQ= 2.34



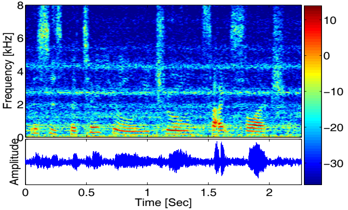
Noisy signal, car interior noise
LSD = 3.48dB, PESQ= 2.47



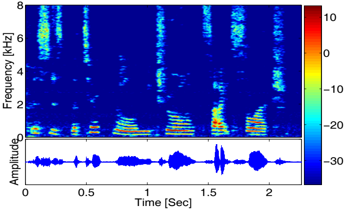
Enhanced signal, OM-LSA
LSD = 2.67dB, PESQ= 3.00



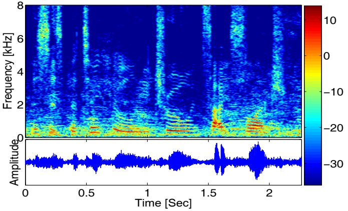
Noisy signal, F16 cockpit noise
LSD = 7.99dB, PESQ= 1.76



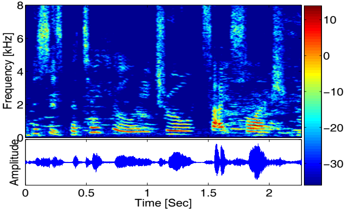
Enhanced signal, OM-LSA
LSD = 4.27dB, PESQ= 2.29



Noisy signal, babble noise
LSD = 5.97dB, PESQ= 1.87



Enhanced signal, OM-LSA
LSD = 4.30dB, PESQ= 2.13



Discussion

- The improvement in SegSNR and reduction in LSD are influenced by the variability of the noise characteristics in time and the initial SegSNR and LSD of the noisy signal.
- The faster the noise spectrum varies in time, the less reliable is the noise spectrum estimator, and consequently the lower is the quality that can be achieved.
- In some applications, a delay of a few short-term frames between the enhanced speech and the noisy observation is tolerable $\Rightarrow$ a noncausal estimation approach may produce less signal distortion and less musical residual noise.

Statistical Models

- Given $\{\lambda_{tk}\}$ and the state of speech presence in each time-frequency bin (H_1^{tk} or H_0^{tk}), the speech spectral coefficients $\{X_{tk}\}$ are generated by

$$X_{tk} = \sqrt{\lambda_{tk}}\, V_{tk}$$

where $\left\{V_{tk} \mid H_0^{tk}\right\}$ are identically zero, and $\left\{V_{tk} \mid H_1^{tk}\right\}$ are statistically independent complex random variables with zero mean, unit variance, and iid real and imaginary parts:

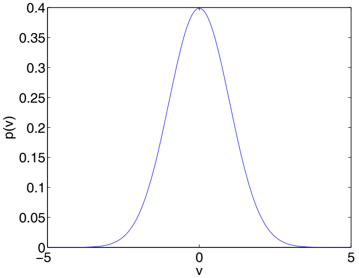
$$\begin{aligned} H_1^{tk} &: \quad E\left\{V_{tk}\right\} = 0\,,\ E\left\{|V_{tk}|^2\right\} = 1 \\ H_0^{tk} &: \quad V_{tk} = 0 \end{aligned}$$

Statistical Models (cont.)

- Gaussian model [McAulay and Malpass, 1980; Ephraim and Malah, 1984]

$$p\left(V_{\rho tk} \mid H_1^{tk} \right) = \frac{1}{\sqrt{\pi}} \exp\left(-V_{\rho tk}^2 \right)$$

$$\rho \in \{R, I\} \ , V_{Rtk} \triangleq \Re\{V_{tk}\} \ , V_{Itk} \triangleq \Im\{V_{tk}\}$$

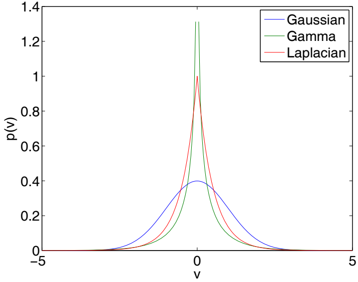


- Gamma model [Porter and Boll, 1984; Martin, 2002]

$$p\left(V_{\rho tk} \mid H_1^{tk}\right) = \frac{1}{2\sqrt{\pi}} \left(\frac{3}{2}\right)^{1/4} \left|V_{\rho tk}\right|^{-1/2} \exp\left(-\sqrt{\frac{3}{2}} \left|V_{\rho tk}\right|\right)$$

- Laplacian model [Martin and Breithaupt, 2003; Lotter and Vary, 2003]

$$p\left(V_{\rho tk} \mid H_1^{tk}\right) = \exp\left(-2\left|V_{\rho tk}\right|\right) .$$



Speech Variance Estimation

The Decision-Directed method

- Over the past two decades, the decision-directed approach has become the acceptable estimation method for variances of speech spectral coefficients [Ephraim and Malah, 1984]

$$\hat{\lambda}_{tk} = \max\left\{\alpha \left|\hat{X}_{t-1,k}\right|^2 + (1-\alpha)\left(\left|Y_{tk}\right|^2 - \sigma_{tk}^2\right) \,, \xi_{\min}\, \sigma_{tk}^2\right\} .$$

Speech Variance Estimation

The Decision-Directed method

- Over the past two decades, the decision-directed approach has become the acceptable estimation method for variances of speech spectral coefficients [Ephraim and Malah, 1984]

$$\hat{\lambda}_{tk} = \max\left\{\alpha\left|\hat{X}_{t-1,k}\right|^2 + (1-\alpha)\left(\left|Y_{tk}\right|^2 - \sigma_{tk}^2\right)\,,\ \xi_{\min}\,\sigma_{tk}^2\right\}\,.$$

- The decision-directed approach is not supported by a statistical model.
- α and $\xi_{\min}$ have to be determined by simulations and subjective listening tests for each particular setup of time-frequency transformation and speech enhancement algorithm.
- α and $\xi_{\min}$ are not adapted to the speech components.

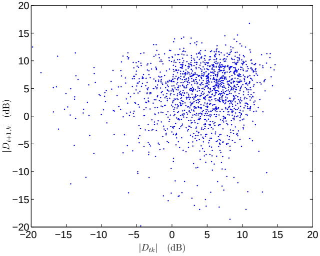
Speech Variance Estimation (cont.)

- Analyze the time-frequency correlation of speech and noise signals in the STFT domain.
- Formulate statistical models for speech signals in the STFT domain, which take into consideration the time-frequency correlation and heavy-tailed distribution of the expansion coefficients.
- Derive estimators for the speech spectral variances, which are based on the proposed models.
- Show that a special case of the proposed variance estimator degenerates to a "decision-directed" estimator.

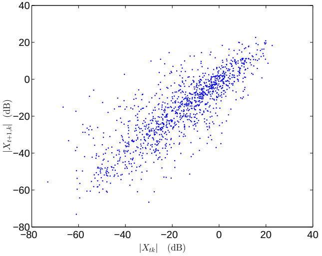
Spectral Analysis

- Clean speech signals, 16 kHz, STFT using Hamming windows, 512 samples length (32 ms), 256 samples framing step (50% overlap).
- Scatter plots for successive spectral magnitudes:

White Gaussian noise



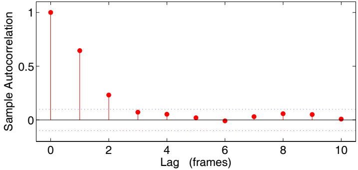
Speech, $k = 17$ (500Hz)



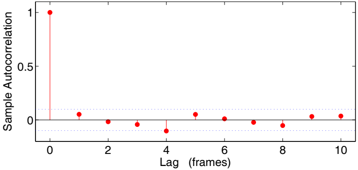
Spectral Analysis (cont.)

- Sample autocorrelation coefficient sequences (ACSs) along time-trajectories:

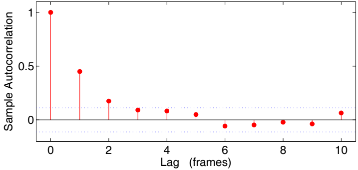
Magnitude, 500Hz, 50% overlap



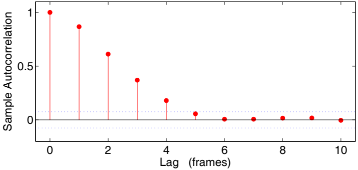
Phase, 500Hz, 50% overlap



Magnitude, 2kHz, 50% overlap

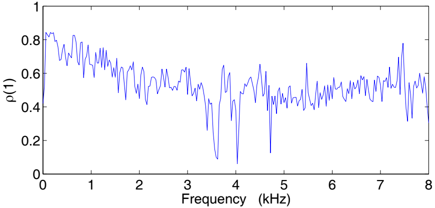


Magnitude, 500Hz, 75% overlap

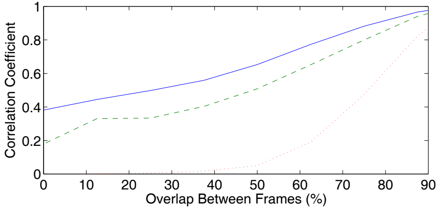


Spectral Analysis (cont.)

- Typical variation of $\rho(1)$, the correlation coefficient between successive spectral magnitudes:



Variation on frequency



Variation on overlap
1kHz (solid)
2kHz (dashed)
noise (dotted)

Discussion

- When observing a time series of successive expansion coefficients in a fixed frequency bin, successive magnitudes of the expansion coefficients are highly correlated, whereas successive phases are nearly uncorrelated.
- Hence, the expansion coefficients are clustered in the sense that large magnitudes tend to follow large magnitudes and small magnitudes tend to follow small magnitudes, while the phase is unpredictable.

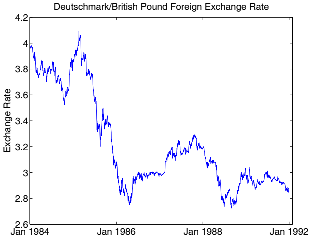
Speech signals in the STFT domain are characterized by volatility clustering and heavy-tailed distribution.

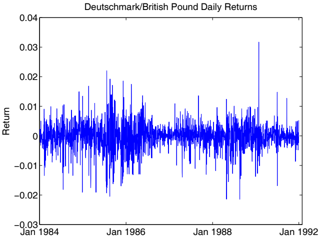
Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Model

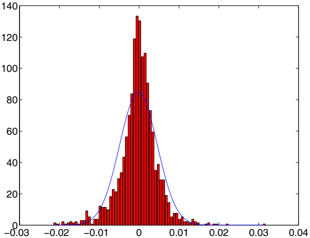
- GARCH models [Engle, 1982; Bollerslev, 1986] are widely used in various financial applications such as risk management, option pricing, and foreign exchange.
- They explicitly parameterize the time-varying volatility in terms of past conditional variances and past squared innovations (prediction errors), while taking into account excess kurtosis (i.e., heavy tail behavior) and volatility clustering, two important characteristics of financial time-series.

$\Rightarrow$ Modeling speech expansion coefficients as GARCH processes offers a reasonable model on which to base the variance estimation, while taking into consideration the heavy-tailed distribution.

GARCH Model (cont.)







GARCH Model (cont.)

Let $\{y_t\}$ denote a real-valued discrete-time stochastic process, and let ψ_t denote the information set available at time t.
The innovation process in the MMSE sense is given by

$$\varepsilon_t = y_t - E\{y_t \mid \psi_{t-1}\}$$

and the conditional variance (volatility) of y_t is defined as

$$\sigma_t^2 = \mathrm{var}\{y_t \mid \psi_{t-1}\} = E\left\{\varepsilon_t^2 \mid \psi_{t-1}\right\} .$$

A GARCH model of order (p, q), denoted by $\varepsilon_t \sim \mathrm{GARCH}(p, q)$, has the following general form

$$\begin{aligned} \varepsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= f\left(\sigma_{t-1}^2, \ldots, \sigma_{t-p}^2, \varepsilon_{t-1}^2 \ldots, \varepsilon_{t-q}^2\right) \end{aligned}$$

where $\{z_t\}$ is a zero-mean unit-variance white noise process with some specified probability distribution.

GARCH Model (cont.)

The widely used GARCH model assumes a linear formulation,

$$\sigma_t^2 = \kappa + \sum_{i=1}^{q} \alpha_i \, \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \, \sigma_{t-j}^2 \,,$$

and the values of the parameters are constrained by

$$\kappa > 0 \,, \; \alpha_i \geq 0 \,, \; \beta_j \geq 0 \,, \quad i = 1, \ldots, q \,, \; j = 1, \ldots, p \,,$$

(sufficient constraints to ensure that the conditional variances $\{\sigma_t^2\}$ are strictly positive) and by

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$$

(necessary and sufficient constraint for the existence of a finite unconditional variance of the innovations process).

Speech Modeling

- As before, given $\{\lambda_{tk}\}$ and the state of speech presence in each time-frequency bin (H_1^{tk} or H_0^{tk}), $\{X_{tk}\}$ are generated by

$$X_{tk} = \sqrt{\lambda_{tk}}\, V_{tk}$$

where $\left\{ V_{tk} \mid H_1^{tk} \right\}$ are statistically independent complex random variables

$$\begin{array}{ll} H_1^{tk}: & E\left\{V_{tk}\right\} = 0\,,\ E\left\{|V_{tk}|^2\right\} = 1 \\ H_0^{tk}: & V_{tk} = 0 \end{array}$$

- However, $\{\lambda_{tk}\}$ are hidden from direct observation even under perfect conditions of zero noise ($D_{tk} = 0$ for all tk).

Speech Modeling (cont.)

- Our approach is to assume that $\{\lambda_{tk}\}$ themselves are random variables, and to introduce *conditional* variances which are estimated from the available information.
- Let $\lambda_{tk|\tau} \triangleq E\left\{|X_{tk}|^2 \mid H_1^{tk}, \mathcal{X}_0^\tau\right\}$ denote the *conditional* variance of X_{tk} under H_1^{tk} given the clean spectral coefficients up to frame τ. We assume that $\lambda_{tk|t-1}$, referred to as the *one-frame-ahead conditional variance*, is a random process which evolves as a $\mathrm{GARCH}(1,1)$ process:

$$\lambda_{tk|t-1} = \lambda_{\min} + \mu\,|X_{t-1,k}|^2 + \delta\left(\lambda_{t-1,k|t-2} - \lambda_{\min}\right)$$

where

$$\lambda_{\min} > 0\,, \quad \mu \geq 0\,, \quad \delta \geq 0\,, \quad \mu + \delta < 1$$

are the standard constraints imposed on the parameters of the GARCH model.

Variance Estimation

Following the rational of Kalman filtering:

- Start with an estimate $\hat{\lambda}_{tk|t-1}$, and update the variance by using the additional information Y_{tk},

Update step:

$$\hat{\lambda}_{tk|t} = E\left\{ |X_{tk}|^2 \,\middle|\, \hat{\lambda}_{tk|t-1}\,,\, Y_{tk} \right\}$$

- Propagate the variance estimate ahead in time to obtain a conditional variance estimate at frame $t+1$,

Propagation step:

$$\hat{\lambda}_{t+1,k|t} = \lambda_{\min} + \mu\,\hat{\lambda}_{tk|t} + \delta\left(\hat{\lambda}_{tk|t-1} - \lambda_{\min}\right)$$

- The propagation and update steps are iterated, to recursively estimate the speech variances as new data arrive.

Relation to Decision-Directed Estimation

- For a Gaussian-GARCH model, the update step can be written as

$$\hat{\lambda}_{tk|t} = \alpha_{tk}\,\hat{\lambda}_{tk|t-1} + (1-\alpha_{tk})\left(|Y_{tk}|^2 - \sigma_{tk}^2\right)$$

with

$$\alpha_{tk} \triangleq 1 - \frac{\hat{\lambda}_{tk|t-1}^2}{\left(\hat{\lambda}_{tk|t-1} + \sigma_{tk}^2\right)^2}\,.$$

- Using the propagation step with $\mu \equiv 1$ and applying the lower bound constraint to $\hat{\lambda}_{tk|t}$ rather than $\hat{\lambda}_{tk|t-1}$, we have

$$\hat{\lambda}_{tk|t} = \max\left\{\alpha_{tk}\,\hat{\lambda}_{t-1,k|t-1} + (1-\alpha_{tk})\left(|Y_{tk}|^2 - \sigma_{tk}^2\right)\,,\ \lambda_{\min}\right\}$$

Relation to Decision-Directed Estimation (cont.)

- Recall the *heuristically motivated* decision-directed estimator [Ephraim and Malah, 1984]

$$\hat{\lambda}_{tk} = \max\left\{\alpha\left|\hat{X}_{t-1,k}\right|^2 + (1-\alpha)\left(\left|Y_{tk}\right|^2 - \sigma_{tk}^2\right)\ ,\ \xi_{\min}\,\sigma_{tk}^2\right\}$$

- A special case of the GARCH-based variance estimator degenerates to a decision-directed estimator with a *time-varying frequency-dependent* weighting factor α_{tk}

$$\alpha \quad \Longleftrightarrow \quad \alpha_{tk}$$

$$\xi_{\min}\,\sigma_{tk}^2 \quad \Longleftrightarrow \quad \lambda_{\min}$$

$$\left|\hat{X}_{t-1,k}\right|^2 \quad \Longleftrightarrow \quad \hat{\lambda}_{t-1,k|t-1} \triangleq E\left\{\left|X_{t-1,k}\right|^2 \,\middle|\, \hat{\lambda}_{t-1,k|t-2}\,,\, Y_{t-1,k}\right\}$$

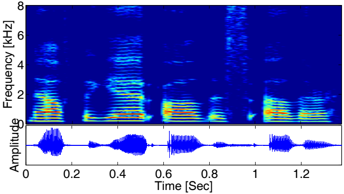
Experimental Results: Setup

- Speech signals: 20 different utterances from 20 different speakers, sampled at 16 kHz and degraded by white Gaussian noise with SNRs in the range $[0, 20]$dB.
- Eight different speech enhancement algorithms are compared

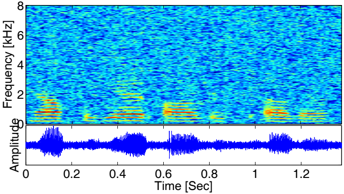
Algorithm #	Statistical Model	Variance Estimation	Fidelity Criterion
1	Gaussian	GARCH	MMSE
2	Gamma	GARCH	MMSE
3	Laplacian	GARCH	MMSE
4	Gaussian	Decision-Directed	MMSE
5	Gamma	Decision-Directed	MMSE
6	Laplacian	Decision-Directed	MMSE
7	Gaussian	GARCH	MMSE-LSA
8	Gaussian	Decision-Directed	MMSE-LSA

Experimental Results

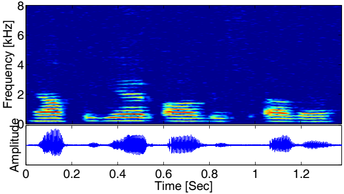
Clean speech signal



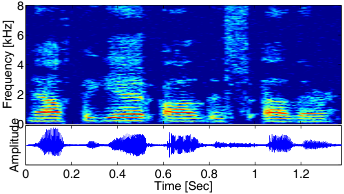
Noisy signal, SNR = 5dB
LSD = 13.75dB, PESQ= 1.76



Decision-Directed, MMSE-LSA
LSD = 9.00dB, PESQ= 2.57



GARCH, MMSE-LSA
LSD = 3.59dB, PESQ= 2.88



Experimental Results (cont.)

Log-Spectral Distortion (LSD)

Input SNR [dB]	GARCH modeling method Gaussian MMSE	GARCH Gaussian LSA	GARCH Gamma MMSE	GARCH Laplacian MMSE	Decision-Directed method Gaussian MMSE	Decision-Directed Gaussian LSA	Decision-Directed Gamma MMSE	Decision-Directed Laplacian MMSE
0	7.77	**4.85**	8.03	7.91	18.89	11.35	17.76	18.14
5	5.78	**4.04**	6.93	6.45	17.29	11.03	15.73	16.26
10	4.14	**3.27**	5.35	4.85	13.87	9.13	11.83	12.48
15	2.50	**2.25**	3.23	2.92	9.19	6.05	6.95	7.59
20	1.30	**1.28**	1.55	1.44	4.88	3.13	2.88	3.34

Perceptual Evaluation of Speech Quality (PESQ) scores (ITU-T P.862)

Input SNR [dB]	GARCH modeling method Gaussian MMSE	GARCH Gaussian LSA	GARCH Gamma MMSE	GARCH Laplacian MMSE	Decision-Directed method Gaussian MMSE	Decision-Directed Gaussian LSA	Decision-Directed Gamma MMSE	Decision-Directed Laplacian MMSE
0	2.52	**2.55**	2.47	2.48	1.91	2.21	1.98	1.96
5	2.97	**2.98**	2.90	2.91	2.30	2.61	2.38	2.36
10	3.37	**3.38**	3.28	3.31	2.70	2.99	2.77	2.75
15	3.67	**3.69**	3.59	3.62	3.09	3.31	3.17	3.15
20	3.88	**3.89**	3.83	3.85	3.53	3.64	3.62	3.60

Conclusions

- The GARCH modeling method yields lower LSD and higher PESQ scores than the decision-directed method.
- Using the decision-directed method, a Gaussian model is inferior to Gamma and Laplacian models.
- Using the GARCH modeling method, a Gaussian model is superior to Gamma and Laplacian models.
- It is difficult, or even impossible, to derive analytical expressions for MMSE-LSA estimators under Gamma or Laplacian models.

The GARCH modeling method facilitates MMSE-LSA estimation, while taking into consideration the heavy-tailed distribution.