Markov-Switching GARCH Models and Applications to Digital Processing of Speech Signals

Ari Abramson

Electrical Engineering Department
Technion - Israel Institute of Technology

Supervised by: Prof. Israel Cohen

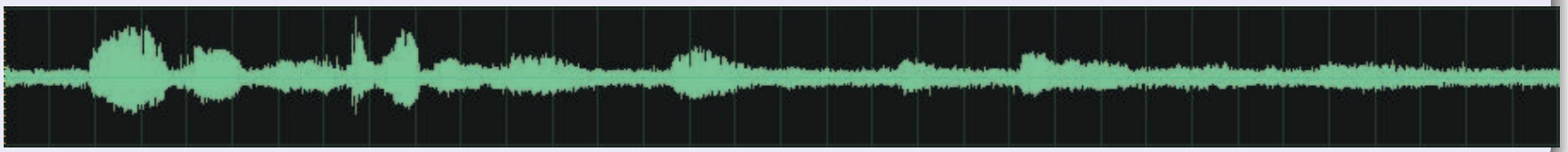
Outline

- Introduction - Spectral Enhancement
- Simultaneous Detection and Estimation for Speech Enhancement
- Markov-Switching GARCH Model
- Experimental Results

Speech Enhancement - Main Goal

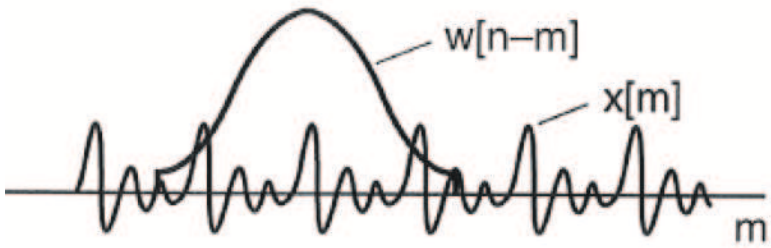
Given a noisy observation of speech signal, $y = x + d$

find the *best* estimate for the speech signal $\hat{x}$

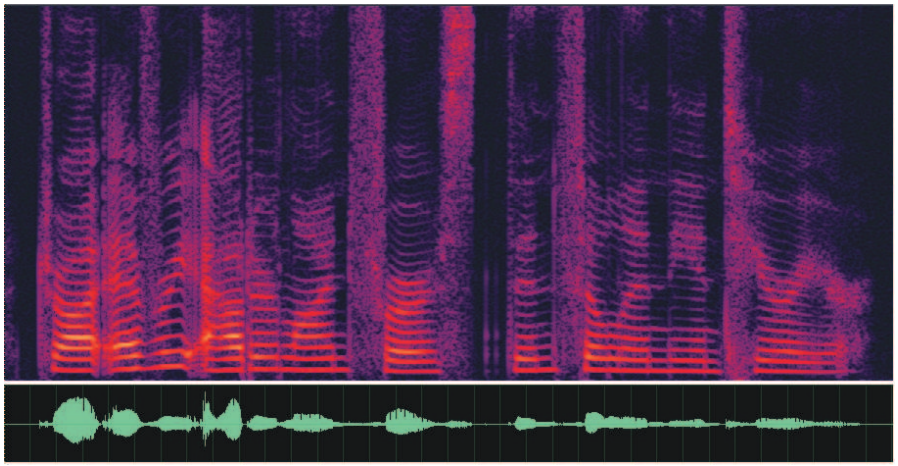


Short-Time Fourier Transform (STFT)

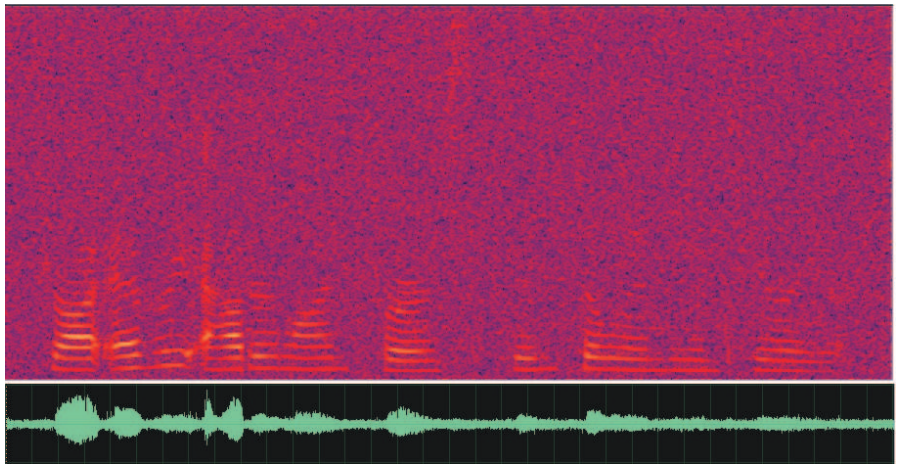
Since the signal is highly non-stationary, STFT is applied to yield a time-frequency representation:



and we get $Y_{tk} = X_{tk} + D_{tk}$



clean speech



noisy speech, SNR=5 dB

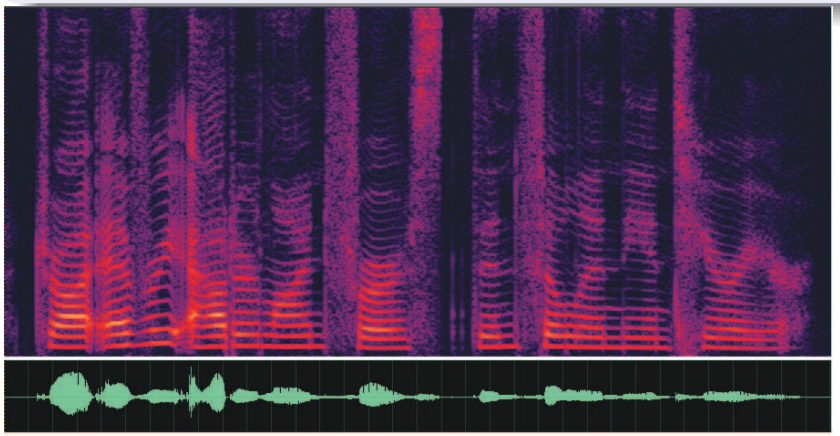
Spectral Enhancement

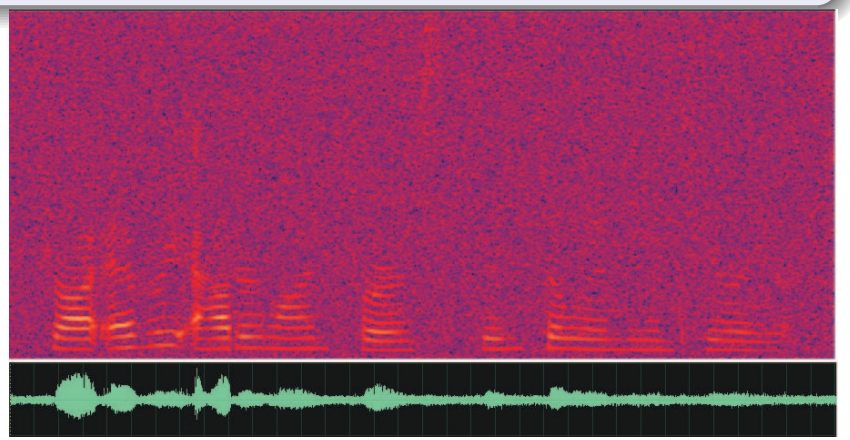
The expansion coefficients are sparse in the STFT domain

$$\begin{aligned} \mathcal{H}_1^{tk}\,(\text{speech present}): \quad Y_{tk} &= X_{tk} + D_{tk} \\ \mathcal{H}_0^{tk}\,(\text{speech absent}): \quad Y_{tk} &= D_{tk} \end{aligned}$$

and the spectral enhancement problem can be formulated as

$$\min_{\hat{X}_{tk}} E\left\{ d\left(X_{tk}, \hat{X}_{tk}\right) \,\middle|\, \hat{q}_{tk}\,, \hat{\lambda}_{tk}\,,\ \hat{\lambda}_{d,tk}\,,\ Y_{tk} \right\}$$





Spectral Enhancement

The expansion coefficients are sparse in the STFT domain

$$\begin{aligned} \mathcal{H}_1^{tk}\,(\text{speech present}): \quad Y_{tk} &= X_{tk} + D_{tk} \\ \mathcal{H}_0^{tk}\,(\text{speech absent}): \quad Y_{tk} &= D_{tk} \end{aligned}$$

and the spectral enhancement problem can be formulated as

$$\min_{\hat{X}_{tk}} E\left\{ d\left(X_{tk}, \hat{X}_{tk}\right) \,\middle|\, \hat{q}_{tk}\,, \hat{\lambda}_{tk}\,, \hat{\lambda}_{d,tk}\,, Y_{tk} \right\}$$

- $d\left(X_{tk}, \hat{X}_{tk}\right) = \left| g(X_{tk}) - g(\hat{X}_{tk}) \right|^2$ - distortion measure
- $q_{tk} = P\left(\mathcal{H}_1^{tk}\right)$ - speech presence probability
- $\lambda_{tk} = E\left\{ |X_{tk}|^2 \,|\, \mathcal{H}_1^{tk} \right\}$ - speech spectral variance
- $\lambda_{d,tk} = E\left\{ |D_{tk}|^2 \right\}$ - noise spectral variance

Spectral Enhancement (cont.)

A spectral enhancement system requires:

Estimation / Detection method:

- Fidelity criterion, i.e., $g(X) = X\,,\ |X|\,,\ log|X|$
- A detector for speech spectral coefficients / speech presence probability

and

Spectral model:

- Statistical models for $\{X_{tk}\}$ and $\{D_{tk}\}$
- Spectral variance estimators $\hat{\lambda}_{tk}$ and $\hat{\lambda}_{d,tk}$

Spectral Models

Decision-directed spectral variance estimator

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

$$0 \leq \mu \leq 1$$

[Ephraim & Malah 84]

Hidden-Markov Model

A vector model:

Hidden Markov chain with mixtures of AR processes.

[Ephraim *et al.* 89, 92]

Spectral Models

Decision-directed spectral variance estimator

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

$$0 \leq \mu \leq 1$$

[Ephraim & Malah 84]

Hidden-Markov Model

A vector model:

Hidden Markov chain with mixtures of AR processes.

[Ephraim *et al.* 89, 92]

Existing Algorithms

- **Spectral subtraction:** detection (power thresholding) and independent estimation (power subtraction).
- **Subspace approach:** detection followed by estimation.
- **MMSE, STSA, OM-LSA (and other):** estimation under uncertainty (e.g., $\hat{X}_{tk} = E\{X_{tk} \mid Y_{tk}, \mathcal{H}_1^{tk}\} p(\mathcal{H}_1^{tk} \mid Y_{tk})$).

Drawbacks:

- Estimation under uncertainty degrades speech quality and prevent sufficient attenuation of noise-only coefficients compared to using an ideal detector.
- Erroneous detection may remove desired speech components or result in an annoying musical noise.

Existing Algorithms

- **Spectral subtraction:** detection (power thresholding) and independent estimation (power subtraction).
- **Subspace approach:** detection followed by estimation.
- **MMSE, STSA, OM-LSA (and other):** estimation under uncertainty (e.g., $\hat{X}_{tk} = E\{X_{tk} \mid Y_{tk}, \mathcal{H}_1^{tk}\} p(\mathcal{H}_1^{tk} \mid Y_{tk})$).

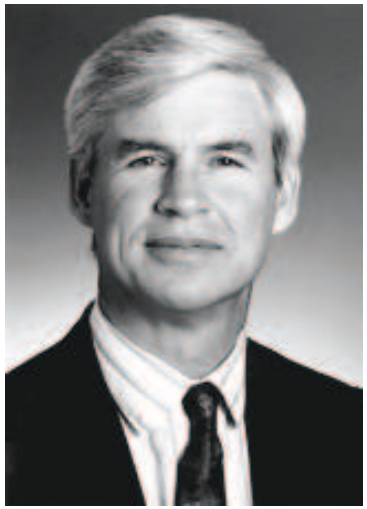
Drawbacks:

- Estimation under uncertainty degrades speech quality and prevent sufficient attenuation of noise-only coefficients compared to using an ideal detector.
- Erroneous detection may remove desired speech components or result in an annoying musical noise.

GARCH:

Generalized Autoregressive Conditional Heteroscedasticity

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Model



Robert F. Engle

Nobel price (2003) in economic sciences "*for methods of analyzing economic time series with time-varying volatility (ARCH)*"

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Model

- GARCH models [Engle '82; Bollerslev '86] are widely used in various financial applications such as risk management, option pricing, and foreign exchange.
- Explicitly parameterize the time-varying volatility in terms of past conditional variances and past squared innovations, while taking into account excess kurtosis and volatility clustering.

Modeling speech expansion coefficients as GARCH processes offers reasonable model on which to base the variance estimation

[Cohen '04, '06]

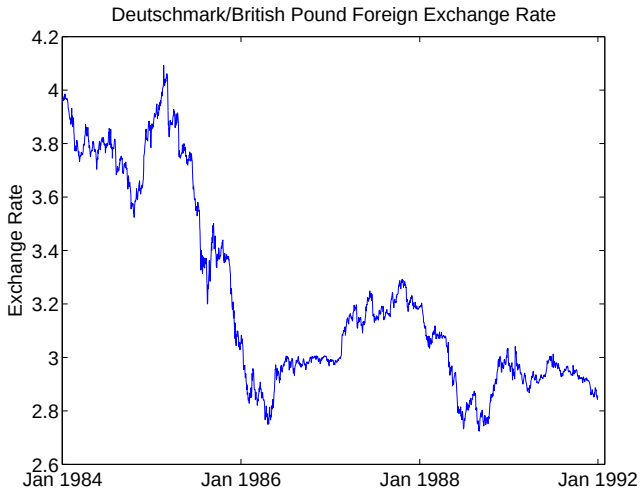
Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Model

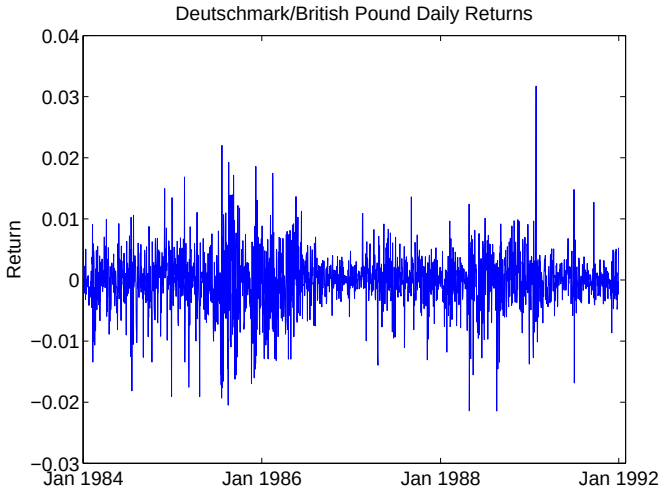
- GARCH models [Engle '82; Bollerslev '86] are widely used in various financial applications such as risk management, option pricing, and foreign exchange.
- Explicitly parameterize the time-varying volatility in terms of past conditional variances and past squared innovations, while taking into account excess kurtosis and volatility clustering.

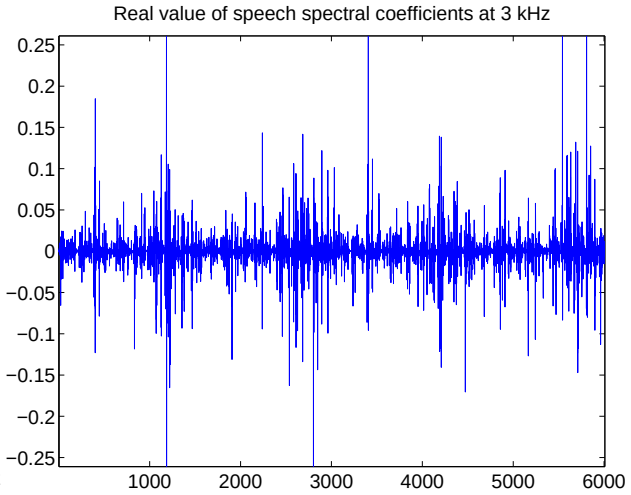
Modeling speech expansion coefficients as GARCH processes offers reasonable model on which to base the variance estimation

[Cohen '04, '06]

GARCH Model (cont.)

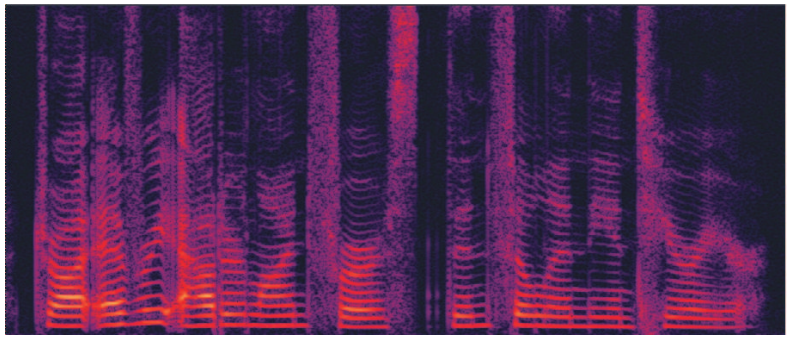


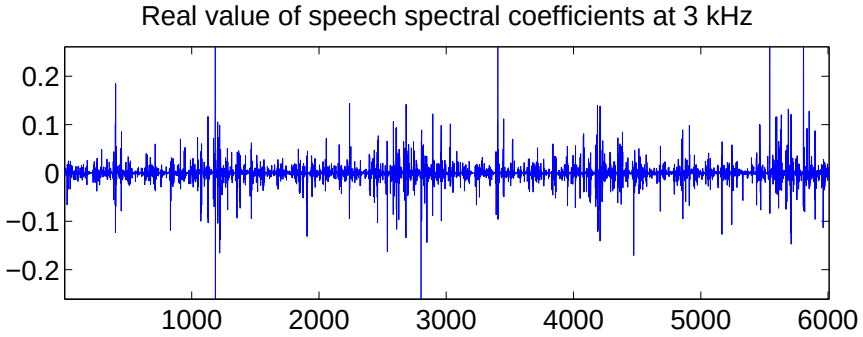




Speech Spectral Analysis

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





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 (prediction error) ε_t at time t in the MMSE sense is given by

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

- The conditional variance (volatility) of y_t is defined by

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

GARCH Model (cont.)

A linear GARCH model of order (p, q), $\varepsilon_t \sim \mathrm{GARCH}(p, q)$:

$$\varepsilon_t = \sigma_t \, v_t$$

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

$\{v_t\}$ is a zero-mean unit-variance white noise process.

To ensure positive conditional variances:

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

and for the existence of a finite unconditional variance:

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

GARCH Model (cont.)

A linear GARCH model of order (p, q), $\varepsilon_t \sim \mathrm{GARCH}(p, q)$:

$$\varepsilon_t = \sigma_t \, v_t$$

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

$\{v_t\}$ is a zero-mean unit-variance white noise process.
To ensure positive conditional variances:

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

and for the existence of a finite unconditional variance:

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

Markov-Switching GARCH Model

- In a Markov-switching GARCH model, the parameters are allowed to change in time according to a Markovian state to enable tracking of any *shocks*.
- Multi-state models are used for speech recognition and for speech enhancement [Drucker '68, Ephraim et al. '89, '92].
- In case of speech signals the parameters may change in time, *e.g.*, for representing different speech phonemes or speakers.
- A special state may be defined for speech absence hypothesis. This may results in a soft voice activity detector.

Markov-Switching GARCH Model

S_t - an m-state, first-order hidden Markov chain with transition probabilities matrix A.

A general (p, q)-order MS-GARCH model follows:

Given that $S_t = s_t$:

$$\begin{aligned} \varepsilon_t &= \sigma_{t,s_t} v_t \\ \sigma^2_{t,s_t} &= \xi_{s_t} + \sum_{i=1}^{q} \alpha_{i,s_t} \varepsilon^2_{t-i} + \sum_{j=1}^{p} \beta_{j,s_t} \sigma^2_{t-j,s_{t-j}} \end{aligned}$$

Asymptotic Stationarity

- For a single-state model: $\sum_i \alpha_i + \sum_j \beta_j < 1$ is *necessary and sufficient* to have an asymptotic stationary variance and to guarantee a finite second-order moment [Bollerslev 86].
- Clearly, for a MS-GARCH model: $\sum_i \alpha_{i,s} + \sum_j \beta_{j,s} < 1 \ \forall s$ is a sufficient condition.
- *Sufficient and necessary* conditions for stationarity are known only for some degenerated cases.

Asymptotic Stationarity (cont.)

The unconditional variance follows:

$$\begin{aligned} E\left[\varepsilon_t^2\right] &= E_{\psi_{t-1},s_t}\left[E\left(\varepsilon_t^2 \mid \psi_{t-1}, s_t\right)\right] \\ &= E_{S_t}\left[E_{\psi_{t-1}}\left(\sigma_{t,s_t}^2 \mid s_t\right)\right] = \sum_{s_t=1}^{m} \pi_{s_t} E_{\psi_{t-1}}\left(\sigma_{t,s_t}^2 \mid s_t\right) \end{aligned}$$

where $\pi_{s_t} \triangleq p\left(S_t = s_t\right)$

Asymptotic Stationarity (cont.)

Define:

$$\psi \triangleq \begin{bmatrix} \mathcal{K}^{(1)} & \mathcal{K}^{(2)} & & \cdots & & \mathcal{K}^{(r)} \\ I_m & 0_m & & \cdots & & 0_m \\ 0_m & I_m & & & & \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ & & & & & \\ 0_m & \cdots & & 0 & I_m & 0_m \end{bmatrix}$$

where

$$\left\{\mathcal{K}^{(i)}\right\}_{s,\tilde{s}} \triangleq \left(\alpha_{i,s} + \beta_{i,s}\right) \frac{\pi_{\tilde{s}}}{\pi_s} \left\{A^i\right\}_{\tilde{s},s}, \quad s, \tilde{s} = 1, ..., m$$

and $r \triangleq \max\{p, q\}$.

Asymptotic Stationarity (cont.)

Let Λ be an $m \times m$ square matrix with $\{\Lambda\}_{ij} = \left\{(I - \Psi)^{-1}\right\}_{ij}$, $i, j = 1, ..., m$. Then:

Theorem

The process is asymptotically wide-sense stationary with asymptotic variance $\lim_{t\to\infty} E\left(\varepsilon_t^2\right) = \boldsymbol{\pi}'\Lambda\boldsymbol{\xi}$, *if and only if* $\rho(\Psi) < 1$

[Abramson and Cohen, Econometric Theory 07]

$\Rightarrow$ Some regimes may allow volatility to grow over time and still the second-order moment will be finite.

Similar method was applied to other MS-GARCH formulations.

Asymptotic Stationarity (cont.)

Let Λ be an $m \times m$ square matrix with $\{\Lambda\}_{ij} = \left\{(I - \Psi)^{-1}\right\}_{ij}$, $i, j = 1, ..., m$. Then:

Theorem

The process is asymptotically wide-sense stationary with asymptotic variance $\lim_{t\to\infty} E\left(\varepsilon_t^2\right) = \boldsymbol{\pi}'\Lambda\boldsymbol{\xi}$, *if and only if* $\rho(\Psi) < 1$

[Abramson and Cohen, Econometric Theory 07]

$\Rightarrow$ Some regimes may allow volatility to grow over time and still the second-order moment will be finite.

Similar method was applied to other MS-GARCH formulations.

Back to Speech:

Spectral Modeling of Speech Signals Using MS-GARCH

MS-GARCH Model in the Time-Frequency Domain

A frame (or subband) dependent Markov chain S_t with realization $s_t \in 0, 1, ..., m$

$X_{tk} \sim MS - GARCH(1,1)$:

- $X_{tk} = \sqrt{\lambda_{tk|t-1,s_t}} V_{tk}$
- $\lambda_{tk|t-1,s_t} = \lambda_{\min,s_t} + \alpha_{s_t} \left|X_{t-1,k}\right|^2 + \beta_{s_t} \left(\lambda_{t-1,k|t-2,s_{t-1}} - \lambda_{\min,s_{t-1}}\right)$

with $\lambda_{\min,s_t} > 0,\ \alpha_{s_t}, \beta_{s_t} \geq 0$. $\{V_{tk}\}$ are iid complex random variables with zero mean and unit variance.

$\Rightarrow$ $\{X_{tk}\}$ are statistically dependent but, $\left\{X_{tk} \,|\, \lambda_{tk|t-1,s_t}, s_t\right\}$ are zero-mean statistically independent random variables.

MS-GARCH Model in the Time-Frequency Domain

A frame (or subband) dependent Markov chain S_t with realization $s_t \in 0, 1, ..., m$

$X_{tk} \sim MS - GARCH(1,1)$:

- $X_{tk} = \sqrt{\lambda_{tk|t-1,s_t}} V_{tk}$
- $\lambda_{tk|t-1,s_t} = \lambda_{\min,s_t} + \alpha_{s_t} \left|X_{t-1,k}\right|^2 + \beta_{s_t} \left(\lambda_{t-1,k|t-2,s_{t-1}} - \lambda_{\min,s_{t-1}}\right)$

with $\lambda_{\min,s_t} > 0,\ \alpha_{s_t}, \beta_{s_t} \geq 0$. $\{V_{tk}\}$ are iid complex random variables with zero mean and unit variance.

$\Rightarrow$ $\{X_{tk}\}$ are statistically dependent but, $\left\{X_{tk} \mid \lambda_{tk|t-1,s_t},\ s_t\right\}$ are zero-mean statistically independent random variables.

Model Estimation

- Parameters are traditionally estimated from a training set using ML approach.
- Alternative, parameters may be chosen such that each state would represent a different level of the spectral energy.
- Setting $\alpha_s = \beta_s = 0$ for a specific s results in a constant variance, hence, may be used for speech absence.

Model estimation (cont.)

- The conditional variances in each subband (indexed n) are limited within a dynamic range of η_g dB.
- For the speech absence state (namely, $s_t = 0$):

$$\lambda_{\min,n,0} = 10^{\log_{10}\zeta_g - \eta_g/10}\,,\ \alpha_{n,0} = \beta_{n,0} = 0\,.$$

  where $\zeta_g \triangleq \max_{t,k} |X_{tk}|^2$.
- Under speech presence, η_t dB is the local dynamic range of the conditional variances such that

$$\lambda_{\min,n,1} = \max\left\{\lambda_{\min,n,0}, 10^{\log_{10}\zeta_n - \eta_t/10}\right\}\,,$$

  and $\lambda_{\min,n,s}\,,\ s = 2, ..., m$ are log-spaced between $\lambda_{\min,n,1}$ and $\zeta_n \triangleq \max_{t,k\in\kappa_n} |X_{tk}|^2$.

Model estimation (cont.)

- The parameters $\alpha_{n,s}, \beta_{n,s}$ for $s > 0$ set the volatility level of the conditional variance.
- Under an immutable state s, the stationary variance follows

$$\lambda_{\infty,n,s} \triangleq \lim_{t\to\infty,\, k\in\kappa_n} \lambda_{tk|t-1,s} = \lambda_{\min,n,s} \frac{1-\beta_{n,s}}{1-\alpha_{n,s}-\beta_{n,s}}$$

  assuming $\alpha_{n,s} + \beta_{n,s} < 1$.
- Different states are related to different dynamic ranges in ascending order, thus $\lambda_{\infty,n,s} \leq \lambda_{\min,n,s+1}$ and therefore

$$\frac{1-\beta_{n,s}}{1-\alpha_{n,s}-\beta_{n,s}} \leq \frac{\lambda_{\min,n,s+1}}{\lambda_{\min,n,s}} .$$

Relation to HMM

- In a standard HMM:

$$p\left(\mathbf{X}_t \,|\, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots, s_t, s_{t-1}, s_{t-2}, \ldots\right) = p\left(\mathbf{X}_t \,|\, s_t\right)$$

$\Rightarrow$ each state specifies a specific pdf.

- In a MS-GARCH process both past observations and the regime-path are required for the conditional density.

$\Rightarrow$ each state formulates the evolution of the conditional variance.

Reconstruction of the Conditional Variance

In a *noiseless* environment (econometric) the conditional variance is recursively reconstructed. In our case, the conditional variance is estimated using two steps: [Abramson and Cohen, Trans. SP '07]

Propagation

$$\hat{\lambda}_{tk|t-1,s_t} = \lambda_{\min,s_t} + \alpha_{s_t} E\left\{ |X_{t-1,k}|^2 \mid \mathcal{Y}^{t-1}, s_t \right\} + \beta_{s_t} E\left\{ \lambda_{t-1,k|t-2,s_{t-1}} \mid \mathcal{Y}^{t-1}, s_t \right\} - \beta_{s_t} E\left\{ \lambda_{\min,s_{t-1}} \mid \mathcal{Y}^{t-1}, s_t \right\}$$

$$E\left\{ |X_{t-1,k}|^2 \mid \mathcal{Y}^{t-1}, s_t \right\} = \sum_{s_{t-1}} p\left(s_{t-1} \mid s_t, \mathcal{Y}^{t-1}\right) E\left\{ |X_{t-1,k}|^2 \mid \mathcal{Y}^{t-1}, s_{t-1} \right\}$$

Reconstruction of the Conditional Variance (cont.)

and

Update

$$\begin{aligned}\hat{\lambda}_{tk|t,s_t} =& E\left\{ |X_{tk}|^2 \,\middle|\, s_t, \hat{\lambda}_{tk|t-1,s_t}, Y_t \right\} \\ =& \frac{\hat{\lambda}_{tk|t-1,s_t}}{\hat{\lambda}_{tk|t-1,s_t} + \sigma_{tk}^2} \left[\sigma_{tk}^2 + \frac{\hat{\lambda}_{tk|t-1,s_t}}{\hat{\lambda}_{tk|t-1,s_t} + \sigma_{tk}^2} |Y_{tk}|^2 \right]\end{aligned}$$

Relation to decision directed estimation

- Recall the decision-directed estimation:

$$\hat{\lambda}_{tk}^{DD} = \max\left\{\mu\left|\hat{X}_{t-1,k}\right|^2 + (1-\mu)\left(\left|Y_{tk}\right|^2 - \sigma_{tk}^2\right), \xi_{\min}\sigma_{tk}^2\right\} \quad ,0 \leq \mu \leq 1$$

- For an $ARCH(1)$ model (*i.e.*, $\beta = 0$) with $\alpha = 1$ the update step can be written as [Abramson & Cohen, Trans. SP]

$$\hat{\lambda}_{tk|t} = \bar{\mu}_{tk} E\left\{\left|X_{t-1,k}\right|^2 \mid \mathcal{Y}^{t-1}\right\} + (1-\bar{\mu}_{tk})\left(\left|Y_{tk}\right|^2 - \sigma_{tk}^2\right) + \bar{\mu}_{tk}\lambda_{\min}$$

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

Relation to decision directed estimation (cont.)

- For $\lambda_{\min} << E\left\{|X_{t-1,k}|^2 \mid \mathcal{Y}^{t-1}\right\}$, the degenerated ARCH-based variance estimation with $\alpha = 1$ is similar to the decision-directed approach with

$$\begin{array}{ccc} \mu & \Longleftrightarrow & \bar{\mu}_{tk} \\ \left|\hat{X}_{t-1,k}\right|^2 & \Longleftrightarrow & E\left\{|X_{t-1,k}|^2 \mid \mathcal{Y}^{t-1}\right\} \end{array}$$

- GARCH modeling allows $\alpha < 1$ and $\beta > 0$ and it refers to the spectral variances as a *random process* rather than a sequence of random parameters which are heuristically evaluated.

State Probability

The conditional state probability is derived by

$$p\left(s_t \,|\, \mathcal{Y}^t\right) = \frac{p\left(\mathbf{Y}_t \,|\, s_t, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right) p\left(s_t \,|\, \mathcal{Y}^{t-1}\right)}{\sum_{s_t} p\left(\mathbf{Y}_t | s_t, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right) p\left(s_t \,|\, \mathcal{Y}^{t-1}\right)},$$

where

$$p\left(s_t \,|\, \mathcal{Y}^{t-1}\right) = \sum_{s_{t-1}} p\left(s_{t-1} \,|\, \mathcal{Y}^{t-1}\right) a_{s_{t-1},s_t}.$$

Accordingly,

$1 - p\left(s_t = 0 \,|\, \mathcal{Y}^t\right)$ is a soft voice activity detector for the subband.

- The problem of state smoothing, $p(s_t \,|\, \mathcal{Y}^{t+L})$ with $L > 0$, is also addressed [Abramson and Cohen, SPL '06].

State smoothing

- State smoothing, *i.e.*, $p\left(s_t|\mathcal{Y}^{\tau}\right)$ for $\tau > t$ in HMP relies on [Chang and Hancock '66,Lindgren '78, Askar and Derin '81]

$$f\left(\mathbf{X}_t \,|\, \psi_{t-1}, s_t\right) = f\left(\mathbf{X}_t \,|\, s_t\right)$$

- In an AR-HMP a finite set of past observation is required [Kim '94].
- In a path-dependent MS-GARCH process the density is a function of all past values and active states.
- We derive generalization of both the *forward-backward recursions* and the *stable backward recursion* to path dependent model [Abramson and Cohen, IEEE Signal processing letters '06].

State smoothing (cont.)

Define

$$\hat{\Lambda}_t \triangleq \left\{ \hat{\boldsymbol{\lambda}}_{t|t-1,S_t} \,|\, S_t = 1, \ldots, m \right\}$$

the *generalized forward density*

$$\alpha\left(s_t, \mathcal{Y}^t\right) \triangleq f\left(s_t, \hat{\Lambda}_t, \mathbf{Y}_t\right)$$

and the *generalized backward density*

$$\beta\left(\mathcal{Y}_{t+l}^{t+L} \,|\, \mathcal{S}_t^{t+l-1}, \mathcal{Y}^{t+l-1}\right) \triangleq f\left(\mathcal{Y}_{t+l}^{t+L} \,|\, \mathcal{S}_t^{t+l-1}, \hat{\Lambda}_t, \mathcal{Y}_t^{t+l-1}\right)$$

then the noncausal state probability can be obtained by

$$p\left(s_t | \mathcal{Y}^{t+L}\right) = p\left(s_t \,|\, \hat{\Lambda}_t, \mathcal{Y}_t^{t+L}\right) = \frac{\alpha\left(s_t, \mathcal{Y}^t\right) \beta\left(\mathcal{Y}_{t+1}^{t+L} \,|\, s_t, \mathcal{Y}^t\right)}{\sum_{s_t} \alpha\left(s_t, \mathcal{Y}^t\right) \beta\left(\mathcal{Y}_{t+1}^{t+L} \,|\, s_t, \mathcal{Y}^t\right)}$$

Generalized forward recursion

The generalized forward density satisfies the following recursion:

$$\alpha\left(s_t, \mathcal{Y}^t\right) = \begin{cases} \pi_{s_0} f\left(\mathbf{Y}_0 \mid s_0\right) & t = 0 \\ f\left(\mathbf{Y}_t \mid s_t, \hat{\boldsymbol{\lambda}}_{t|t,s_t}\right) \sum_{s_{t-1}} \alpha\left(s_{t-1}, \mathcal{Y}^{t-1}\right) a_{s_{t-1}s_t} & t = 1, 2 \ldots \end{cases}$$

Generalized backward recursion

The generalized backward density requires two recursive steps:

Step I: for $l = 1, ..., L$, for all $\mathcal{S}_t^{t+l}$:

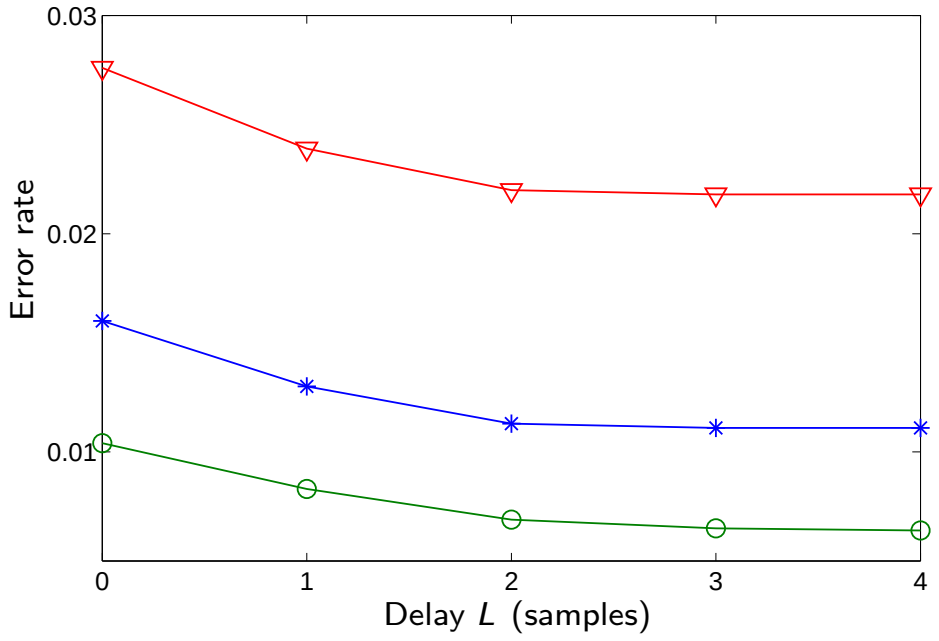
$$\begin{cases} \hat{\boldsymbol{\lambda}}_{t+l|t+l-1,\mathcal{S}_t^{t+l}} = \xi_{s_{t+l}}\mathbf{1} + \alpha_{s_{t+l}}\hat{\boldsymbol{\lambda}}_{t+l-1|t+l-1,\mathcal{S}_t^{t+l-1}} + \beta_{s_{t+1}}\hat{\boldsymbol{\lambda}}_{t+l-1|t+l-2,\mathcal{S}_t^{t+l-1}} \\ \hat{\boldsymbol{\lambda}}_{t+l|t+l,\mathcal{S}_t^{t+l}} = g\left(\hat{\boldsymbol{\lambda}}_{t+l|t+l-1,\mathcal{S}_t^{t+l}}, \mathbf{Y}_{t+l}, \boldsymbol{\sigma}^2\right) \end{cases}$$

Step II: for $l = L, ..., 1$, for all $\mathcal{S}_t^{t+l}$:

$$\begin{cases} f\left(\mathcal{Y}_{t+l}^{t+L} \,|\, \mathcal{S}_t^{t+l}, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t}, \mathcal{Y}_t^{t+l-1}\right) = \beta\left(\mathcal{Y}_{t+l+1}^{t+L} \,|\, \mathcal{S}_t^{t+l}, \mathcal{Y}^{t+l}\right) f\left(\mathbf{Y}_{t+l} \,|\, \mathcal{S}_t^{t+l}, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t}, \mathcal{Y}_t^{t+l-1}\right) \\ \beta\left(\mathcal{Y}_{t+l}^{t+L} \,|\, \mathcal{S}_t^{t+l-1}, \mathcal{Y}^{t+l-1}\right) = \sum_{s_{t+l}} f\left(\mathcal{Y}_{t+l}^{t+L} \,|\, \mathcal{S}_t^{t+l}, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t}, \mathcal{Y}_t^{t+l-1}\right) a_{s_{t+l-1}s_{t+l}} \end{cases}$$

State smoothing - results

- The generalized recursions capture the non-memoryless structure of the Markov-switching GARCH model.
- Each step of the *generalized backward recursion* is calculated for m^{L+1} regime sequences.



State smoothing error rate for 3-state MSTF-GARCH models with SNRs of 5 dB (triangle), 10 dB (asterisk) and 15 dB (circle).

Spectral enhancement

Minimizing

$$E\left\{\left|g\left(X_{tk}\right)-g(\hat{X}_{tk})\right|^2 \mid \mathcal{Y}^t\right\}$$

yields

$$g\left(\hat{X}_{tk}\right)=E\left\{g\left(X_{tk}\right) \mid \mathcal{Y}^t\right\}$$

where

$$E\left\{g\left(X_{tk}\right) \mid \mathcal{Y}^t\right\}=\sum_{s_t} p\left(S_t=s_t \mid \mathcal{Y}^t\right) E\left\{g\left(X_{tk}\right) \mid s_t, \mathcal{Y}^t\right\}$$

Spectral Enhancement

Having the set $\left\{\hat{\lambda}_{tk|t,s_t}\right\}_{s_t=0}^{s_t=m}$, minimizing the mean-square error of the log-spectral amplitude (LSA)

$$E\left\{\left(\log|X_{tk}| - \log\left|\hat{X}_{tk}\right|\right)^2 \,|\, \mathcal{Y}^t\right\}$$

yields:

$$\hat{X}_{tk} = Y_{tk} \prod_{s_t} G_{LSA}\left(\hat{\xi}_{tk,s_t}, \hat{\gamma}_{tk,s_t}\right)^{p(s_t \,|\, \mathcal{Y}^t)},$$

where $G_{LSA}(\xi, \vartheta)$ is the LSA gain function [Ephraim and Malah '85].

Simultaneous Detection and Estimation

Simultaneous Detection and Estimation

Recall

Find an estimate $\hat{X}_{tk}$ which minimizes

$$E\left\{\left|g\left(X_{tk}\right)-\tilde{g}\left(\hat{X}_{tk}\right)\right|^{2} \mid \psi_{t}\right\}$$

where $g\left(X\right)$ and $\tilde{g}\left(X\right)$ are specific functions, *e.g.*, X, $|X|$ or $\log|X|$.

Alternatively,

Bayesian formulation:

$$argmin_{\hat{X}_{tk}} \int C\left(X_{tk},\hat{X}_{tk}\right) p\left(Y_{tk} \mid X_{tk}\right) p(X_{tk})\, dX_{tk}$$

where $C\left(X,\hat{X}\right)=\left|g\left(X\right)-\tilde{g}\left(\hat{X}\right)\right|^{2}$.

Simultaneous Detection and Estimation

Recall

Find an estimate $\hat{X}_{tk}$ which minimizes

$$E\left\{\left|g\left(X_{tk}\right)-\tilde{g}\left(\hat{X}_{tk}\right)\right|^{2} \mid \psi_{t}\right\}$$

where $g\left(X\right)$ and $\tilde{g}\left(X\right)$ are specific functions, *e.g.*, X, $|X|$ or $\log|X|$.

Alternatively,

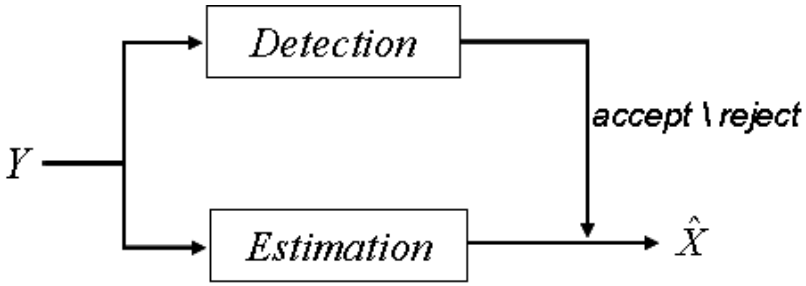
Bayesian formulation:

$$argmin_{\hat{X}_{tk}} \int C\left(X_{tk},\hat{X}_{tk}\right) p\left(Y_{tk} \mid X_{tk}\right) p(X_{tk})\, dX_{tk}$$

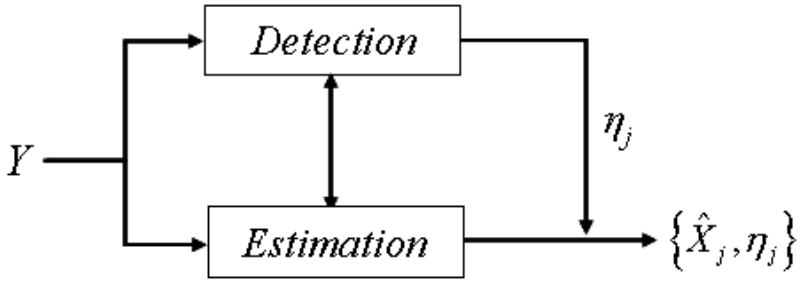
where $C\left(X,\hat{X}\right)=\left|g\left(X\right)-\tilde{g}\left(\hat{X}\right)\right|^{2}$.

Simultaneous Detection and Estimation (cont.)

Independent detection and estimation:



Simultaneous detection and estimation:



Simultaneous Detection and Estimation (cont.)

Guidelines [Middleton *et al.* '68, '72]:

- Integrated detector with a decision space $\left\{\eta_0^{tk}, \eta_1^{tk}\right\}$.
- Under η_j^{tk}, H_j^{tk} is accepted and $\hat{X}_{tk} = \hat{X}_{tk,j}$ is considered.
- Cost for the hypothesis-decision pair $\{H_i, \eta_j\}$:

$$C_{ij}\left(X, \hat{X}\right) = b_{ij}\, d_{ij}\left(X, \hat{X}\right)$$

  - $d_{ij}\left(X, \hat{X}\right)$ is an appropriate distortion measure between the signal and its estimate under η_j.
  - b_{ij} is a trade-off parameter of the cost associated with the pair $\{H_i, \eta_j\}$ against other pairs.

Simultaneous Detection and Estimation (cont.)

- For the signal-observation $\{X, Y\}$ the cost associated with the decision η_j is

$$\begin{aligned} C_j(X,Y) &= \sum_{i=0}^{1} C_{ij}\left(X,\hat{X}\right) p(H_i \mid X) \\ &= \sum_{i=0}^{1} p(H_i)\, C_{ij}\left(X,\hat{X}\right) p(X \mid H_i)/p(X) \end{aligned}$$

- The Combined Bayes risk of simultaneous detection and estimation:

$$R = \sum_{j=0}^{1} \int_Y \int_X C_j(X,Y)\, p(\eta_j \mid Y)\, p(Y \mid X)\, p(X)\; dXdY$$

Simultaneous Detection and Estimation (cont.)

The goal:

Find the optimal detector and estimator which minimize the combined Bayes risk

$$R = \sum_{j=0}^{1} \int_{Y} p\left(\eta_j \mid Y\right) \sum_{i=0}^{1} \int_{X} C_{ij}\left(X, \hat{X}\right) p\left(H_i\right) p\left(Y \mid X\right) p\left(X \mid H_i\right) \, dX dY$$

Simultaneous Detection and Estimation (cont.)

Combined solution:

Optimal estimator under a decision η_j:

$$\hat{X}_j = \arg\min_{\hat{X}} \sum_{i=0}^{1} p\left(H_i\right) r_{ij}\left(Y\right)$$

Optimal decision rule:

$$q\left[r_{10}\left(Y\right) - r_{11}\left(Y\right)\right] \underset{\eta_0}{\overset{\eta_1}{\gtrless}} \left(1-q\right)\left[r_{01}\left(Y\right) - r_{00}\left(Y\right)\right]$$

$r_{ij}\left(Y\right) \triangleq \int C_{ij}\left(X, \hat{X}\right) p\left(Y \mid X\right) p\left(X \mid H_i\right) dX$

$q \triangleq p(H_1)$

Simultaneous Detection and Estimation (cont.)

- Recall
$$C_{ij}\left(X,\hat{X}\right) = b_{ij}\, d_{ij}\left(X,\hat{X}\right)$$
- Normalize $b_{00} = b_{11} = 1$.
- b_{10} is associated with missed-detection, b_{01} is associated with false-detection.
- $d_{ii}(\cdot \mid \cdot)$ is not necessarily zero (estimation error).
- Under H_0, a natural background noise level is desired.

Distortion Functions:

Quadratic distortion measure:

$$d_{ij}\left(X,\hat{X}_j\right)=\begin{cases}\left|X-\hat{X}_j\right|^2 & i=1\\ \left|G_f Y-\hat{X}_j\right|^2 & i=0\end{cases}$$

Quadratic spectral amplitude (QSA) distortion measure:

$$d_{ij}\left(X,\hat{X}_j\right)=\begin{cases}\left(|X|-\left|\hat{X}_j\right|\right)^2 & i=1\\ \left(G_f\,|Y|-\left|\hat{X}_j\right|\right)^2 & i=0\end{cases}$$

$G_f << 1$ is a constant attenuation floor.

Simultaneous Detection and Estimation (cont.)

Definitions:

- *a priori* and *a posteriori* SNRs

$$\xi \triangleq \frac{\lambda}{\lambda_d}, \qquad \gamma \triangleq \frac{|Y|^2}{\lambda_d}, \qquad \upsilon \triangleq \frac{\xi\gamma}{1+\xi}.$$

- Generalized likelihood ratio:

$$\Lambda(\xi,\gamma) = \frac{q}{(1-q)} \frac{p(Y \mid H_1)}{p(Y \mid H_0)}.$$

- $\phi_j(\xi,\gamma) \triangleq b_{0j} + b_{1j}\Lambda(\xi,\gamma)$.

QSA distortion measure:

Under a Gaussian model: [Abramson and Cohen, Trans. ASLP 07]

Estimator

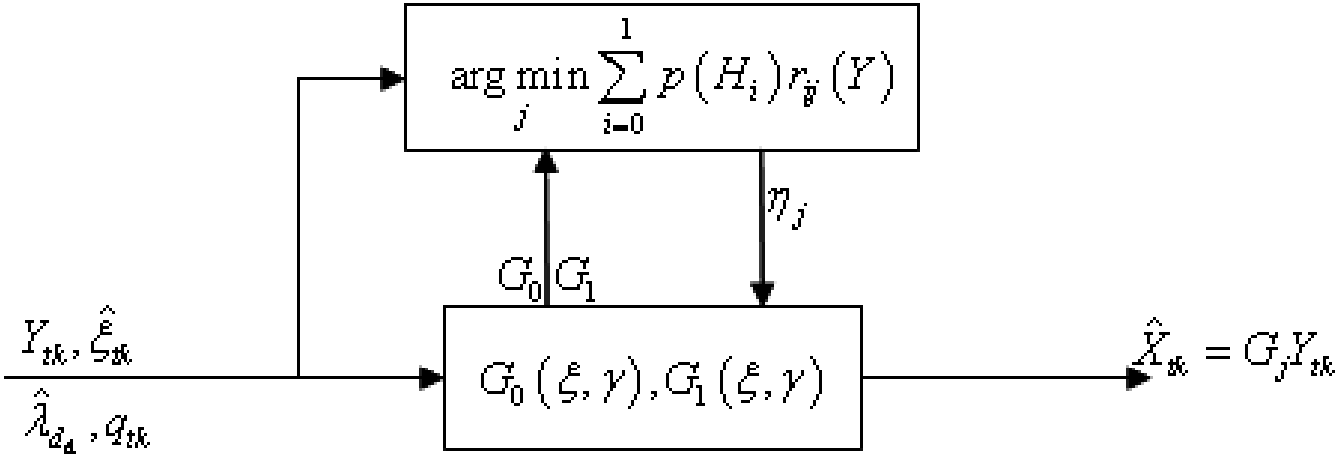
$$\hat{X}_j = \frac{b_{1j}\Lambda(\xi,\gamma)\, G_{STSA}(\xi,\gamma) + b_{0j}\, G_f}{\phi_j(\xi,\gamma)} Y \triangleq G_j(\xi,\gamma)\, Y\,, \qquad j = 0,1$$

Detector

$$\frac{q\, e^{\upsilon}}{1+\xi}\left\{ b_{10} G_0^2 - G_1^2 + \frac{\xi}{(1+\xi)\gamma}(1+\upsilon)(b_{10}-1) + 2\left(G_1 - b_{10} G_0\right) G_{STSA} \right\} \underset{\eta_0}{\overset{\eta_1}{\gtrless}} (1-q)\left[b_{01}\left(G_1 - G_f\right)^2 - \left(G_0 - G_f\right)^2 \right]$$

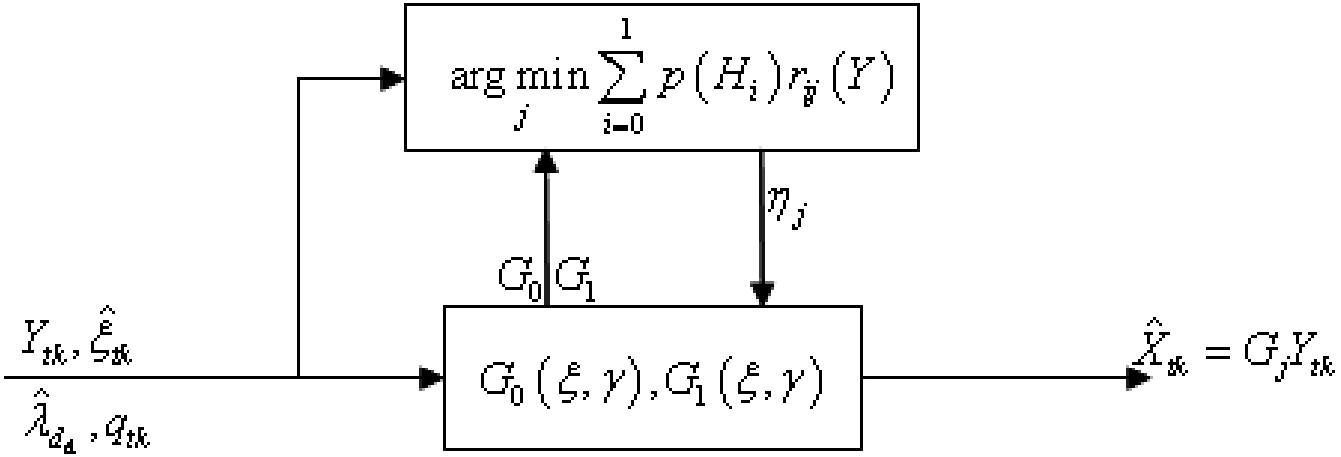
where $G_{STSA} = \frac{\sqrt{2\upsilon}}{\gamma} exp\left(\frac{-\upsilon}{2}\right)\left[(1+\upsilon) I_0\left(\frac{\upsilon}{2}\right) + \upsilon I_1\left(\frac{\upsilon}{2}\right)\right]$ [Ephraim and Malah '84]

Simultaneous Detection and Estimation of Speech Signals:



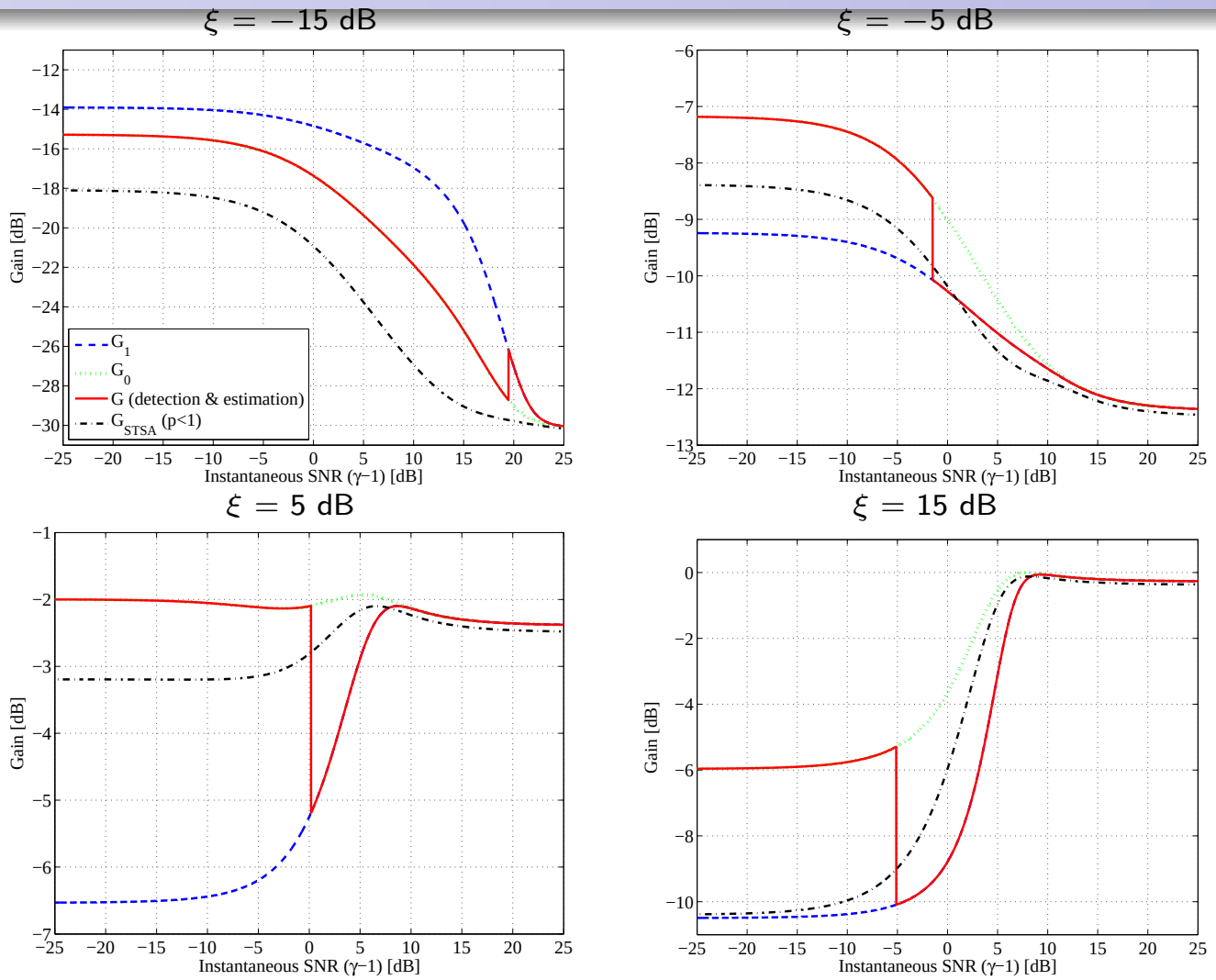
The estimator $\hat{X}_{tk} = G_j Y_{tk}$ is optimal for any given detector.

Simultaneous Detection and Estimation of Speech Signals:



The estimator $\hat{X}_{tk} = G_j Y_{tk}$ is optimal for any given detector.

QSA distortion measure:



$q = 0.8$, $G_f = -20$ dB and $b_{10} = 1.1$, $b_{01} = 5$

Asymptotic behavior

- If the *false alarm* parameter $b_{01} << \Lambda(\xi, \gamma)$ we get $G_1(\xi, \gamma) \simeq G_{STSA}(\xi, \gamma)$.
- If $b_{01} >> \Lambda(\xi, \gamma)$ (wrong decision might be expensive) $\Longrightarrow$ $G_1(\xi, \gamma) \simeq G_f$.
- If the *missed-detection* parameter $b_{10} << \Lambda(\xi, \gamma)^{-1}$ then $G_0(\xi, \gamma) \simeq G_f$.
- But $b_{10} >> \Lambda(\xi, \gamma)^{-1} \Longrightarrow G_0(\xi, \gamma) \simeq G_{STSA}(\xi, \gamma)$ to overcome the high cost for missed-detection.

Relation to classical algorithms

- Recall that $\frac{\Lambda(\xi,\gamma)}{1+\Lambda(\xi,\gamma)} = p\left(H_1 \mid Y\right)$ is the *a posteriori* speech presence probability.
- For equal cost parameters $b_{ij} = 1 \; \forall i, j$ we have

$$\hat{X}_1 = \hat{X}_0 = \left[p\left(H_1 \mid Y\right) G_{STSA}\left(\xi, \gamma\right) + \left(1 - p\left(H_1 \mid Y\right)\right) G_f\right] Y$$

  this is the estimator proposed by [Cohen '06].
- In case $G_f = 0$ we get the STSA suppression rule [Ephraim and Malah '84].

Simultaneous Detection and Estimation

- Design of coupled detector and estimator with cost parameters which control the tradeoff between musical residual noise and speech distortion.
- This method enables to incorporate an independent detector for speech or transient noise with the optimal estimator to yield a suboptimal solution.
- Low computational complexity.

Blind Source Separation

Problem formulation

- Given a single mixture $x = s_1 + s_2$ find estimates $\hat{s}_1$ and $\hat{s}_2$.
- Since the problem is il-posed, some a priori information is required such as statistical *codebook*.
- In the STFT domain, at some specific time-index we have $\mathbf{x} = \mathbf{s}_1 + \mathbf{s}_2$

Problem formulation (cont.)

- $q_1 \in \{1, ..., m_1\}$ and $q_2 \in \{1, ..., m_2\}$ denote the active states of the codebooks with known a priori probabilities $p_1(i) \triangleq p(q_1 = i)$ and $p_2(i) \triangleq p(q_2 = j)$.
- Given that $q_1 = i$ and $q_2 = j$ we assume $\mathbf{s}_1 \sim \mathcal{CN}\left(0, \Sigma_1^{(i)}\right)$ and $\mathbf{s}_2 \sim \mathcal{CN}\left(0, \Sigma_2^{(j)}\right)$.
- Hence, we need to find the active states at each time-frame and subsequently estimate each signal.

Existing solutions

Sequential classification and estimation:

MAP classification:

$$\left\{\hat{i},\hat{j}\right\} = \arg\max_{i,j} p\left(\mathbf{x} \,|\, i,j\right) p\left(i,j\right)$$

MMSE estimation:

$$\hat{\mathbf{s}}_1 = E\left\{\mathbf{s}_1 \,|\, \mathbf{x},\hat{i},\hat{j}\right\} = \Sigma_1^{(\hat{i})}\left(\Sigma_1^{(\hat{i})} + \Sigma_2^{(\hat{j})}\right)^{-1}\mathbf{x} \triangleq W_{\hat{i}\hat{j}}\,\mathbf{x}$$

- Classification error may significantly distort the signal or result in residual interference.

Existing solutions (cont.)

Direct estimation:

$$\begin{aligned}\hat{\mathbf{s}}_1 &= E\left\{\mathbf{s}_1 \,|\, \mathbf{x}\right\} \\ &= \sum_{i,j} p\left(i,j \,|\, \mathbf{x}\right) W_{ij}\, \mathbf{x}\,.\end{aligned}$$

Simultaneous classification and estimation

- Consider a classifier η_{ij} and a combined cost function $C_{ij}^{\overline{ij}}(\mathbf{s}, \hat{\mathbf{s}}) \triangleq b_{ij}^{\overline{ij}} \|\mathbf{s} - \hat{\mathbf{s}}\|_2^2$.
- $b_{ij}^{\overline{ij}} > 0$ are cost parameters for a decision that $\{i, j\}$ is the active pair while $q_1 = \overline{i}$ and $q_2 = \overline{j}$.
- The combined risk:

$$R = \sum_{i,j} \sum_{\overline{i},\overline{j}} \int \int C_{ij}^{\overline{ij}}(\mathbf{s}_1, \hat{\mathbf{s}}_1)\, p\left(\mathbf{x} \,|\, \mathbf{s}_1, \overline{i}, \overline{j}\right) p\left(\mathbf{s}_1 \,|\, \overline{i}, \overline{j}\right) p\left(\overline{i}, \overline{j}\right) p\left(\eta_{ij} \,|\, \mathbf{x}\right) d\mathbf{s}_1\, d\mathbf{x}$$

Simultaneous classification and estimation (cont.)

Using

$$r_{ij}^{\bar{i}\bar{j}}\left(\mathbf{x},\hat{\mathbf{s}}_1\right)=\int C_{ij}^{\bar{i}\bar{j}}\left(\mathbf{s}_1,\hat{\mathbf{s}}_1\right)p\left(\mathbf{x}\,|\,\mathbf{s}_1,\bar{j}\right)p\left(\mathbf{s}_1\,|\,\bar{i}\right)d\mathbf{s}_1$$

we can write:

$$R=\sum_{i,j}\int p\left(\eta_{ij}\,|\,\mathbf{x}\right)\sum_{\bar{i},\bar{j}}p\left(\bar{i},\bar{j}\right)r_{ij}^{\bar{i}\bar{j}}\left(\mathbf{x},\hat{\mathbf{s}}_1\right)d\mathbf{x}$$

and the combined solution is obtained by minimizing:

$$\min_{\eta_{ij},\hat{\mathbf{s}}_1}\left\{R\right\}$$

Optimal solution

The optimal combined solution:

The estimator under the decision η_{ij}

$$\begin{aligned}\hat{\mathbf{s}}_{1,ij} &= \arg\min_{\hat{\mathbf{s}}_1} \sum_{\bar{i},\bar{j}} p\left(\bar{i},\bar{j}\right) r_{ij}^{\bar{i}\bar{j}}\left(\mathbf{x},\hat{\mathbf{s}}_1\right) \\ &= \frac{\sum_{\bar{i}\bar{j}} b_{ij}^{\bar{i}\bar{j}}\, p\left(\mathbf{x}\,|\,\bar{i},\bar{j}\right) p\left(\bar{i},\bar{j}\right) W_{\bar{i}\bar{j}}\,\mathbf{x}}{\sum_{\bar{i}\bar{j}} b_{ij}^{\bar{i}\bar{j}}\, p\left(\mathbf{x}\,|\,\bar{i},\bar{j}\right) p\left(\bar{i},\bar{j}\right)} \\ &\triangleq G_{ij}\,\mathbf{x} \qquad (1)\end{aligned}$$

Optimal solution (cont.)

The optimal classifier is obtained by

$$\min_{ij} \sum_{\bar{i},\bar{j}} p\left(\bar{i},\bar{j}\right) r_{ij}^{\bar{i}\bar{j}}\left(\mathbf{x},\hat{\mathbf{s}}_1\right)$$

where

$$r_{ij}^{\bar{i}\bar{j}}\left(\mathbf{x},\hat{\mathbf{s}}_1\right) = b_{ij}^{\bar{i}\bar{j}}\, p\left(\mathbf{x}\,|\,\bar{i},\bar{j}\right)\left[\mathbf{x}^H\left(W_{\bar{i}\bar{j}}^2 - 2W_{\bar{i}\bar{j}}G_{ij}\right)\mathbf{x} + \mathbf{1}^T\Sigma_2^{(\bar{j})}W_{\bar{i}\bar{j}}\mathbf{1}\right]$$

Parameter selection

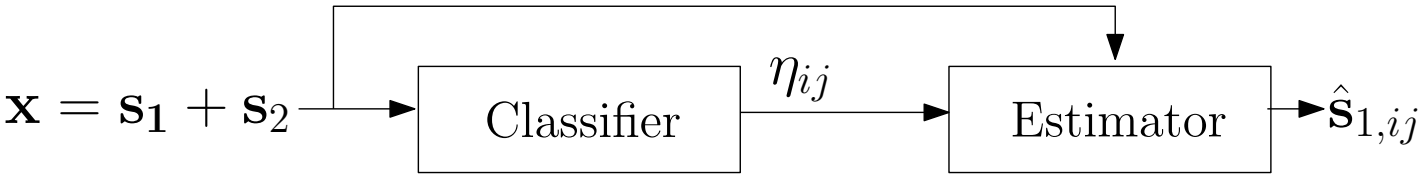
- A specific state is defined for signal absence, e.g., $q = 0$.
- We may choose $b_{ij}^{\bar{i}\bar{j}} = b_i^{\bar{i}} \, b_j^{\bar{j}}$ and

$$b_i^{\bar{i}} = \begin{cases} b_{1,\mathrm{m}} & i = 0, \, \bar{i} \neq 0 \\ b_{1,\mathrm{f}} & i \neq 0, \, \bar{i} = 0 \\ 1 & \text{o.w.} \end{cases}$$

with relation to missed and false detection.

Joint classification and estimation

- Now we consider a cascade classification and estimation:



- Under *signal-presence* decision we wish to control the *residual interference* and under *signal-absence* decision we limit the *distortion level*.

Joint classification and estimation (cont.)

Under signal-presence decision (η_{1j}):

$$\hat{\mathbf{s}}_{1,1j} = \arg\min_{\hat{\mathbf{s}}_1} p\left(q_1 = 1 \,|\, \mathbf{x}\right) E\left\{\|\mathbf{s}_1 - \hat{\mathbf{s}}_1\|_2^2 \,|\, q_1 = 1, \mathbf{x}\right\}$$

$$\text{s.t.} \quad \bar{\varepsilon}_r^2(\mathbf{x}) \leq \sigma_r^2$$

where

$$\bar{\varepsilon}_r^2(\mathbf{x}) \triangleq p\left(q_1 = 0 \,|\, \mathbf{x}\right) E\left\{\|\mathbf{s}_1 - \hat{\mathbf{s}}_1\|_2^2 \,|\, q_1 = 0, \mathbf{x}\right\}$$

Joint classification and estimation (cont.)

Under signal-absence decision (η_{0j}):

$$\hat{\mathbf{s}}_{1,0j} = \arg\min_{\hat{\mathbf{s}}_1} p\left(q_1 = 0 \,|\, \mathbf{x}\right) E\left\{\|\mathbf{s}_1 - \hat{\mathbf{s}}_1\|_2^2 \,|\, q_1 = 0, \mathbf{x}\right\}$$

$$\text{s.t.} \quad \bar{\varepsilon}_d^2(\mathbf{x}) \leq \sigma_d^2 \,,$$

where

$$\bar{\varepsilon}_d^2(\mathbf{x}) \triangleq p\left(q_1 = 1 \,|\, \mathbf{x}\right) E\left\{\|\mathbf{s}_1 - \hat{\mathbf{s}}_1\|_2^2 \,|\, q_1 = 1, \mathbf{x}\right\}$$

Joint classification and estimation (cont.)

The Lagrangian is given by (for η_{0j}):

$$L_d\left(\hat{\mathbf{s}}_1, \mu_d\right) = p\left(q_1 = 0 \,|\, \mathbf{x}\right) E\left\{\|\mathbf{s}_1 - \hat{\mathbf{s}}_1\|_2^2 \mid q_1 = 0, \mathbf{x}\right\} + \mu_d\left(\bar{\varepsilon}_d^2(\mathbf{x}) - \sigma_d^2\right)$$

and

$$\mu_d\left(\bar{\varepsilon}_d^2(\mathbf{x}) - \sigma_d^2\right) = 0 \quad \text{for} \quad \mu_d \geq 0$$

Similarly, under η_{1j} we have $L_r\left(\hat{\mathbf{s}}_1, \mu_r\right)$ with μ_r.

Joint classification and estimation (cont.)

By joining the solutions for both cases we obtain:

$$\hat{\mathbf{s}}_{1,ij} = \frac{\sum_{\bar{i}\bar{j}} \mu_i^{\bar{i}} p\left(\bar{i},\bar{j} \,|\, \mathbf{x}\right) W_{\bar{i}\bar{j}} \mathbf{x}}{\sum_{\bar{i}\bar{j}} \mu_i^{\bar{i}} p\left(\bar{i},\bar{j} \,|\, \mathbf{x}\right)}$$

with

$$\mu_i^{\bar{i}} = \begin{cases} \mu_d & i = 0,\ \bar{i} = 1 \\ \mu_r & i = 1,\ \bar{i} = 0 \\ 1 & \text{o.w.} \end{cases}$$

which is similar to the solution for the simultaneous classification and estimation with $b_j^{\bar{j}} = 1$, $b_{1,\mathrm{m}} = \mu_d$, and $b_{1,\mathrm{f}} = \mu_r$.

GARCH-based codebook

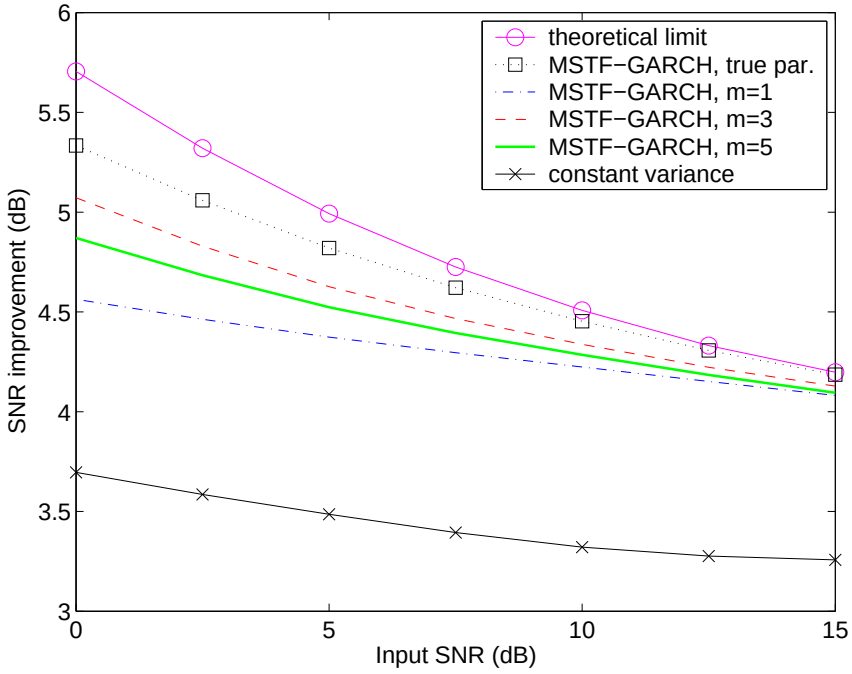
Block diagram

Distortion vs. interference reduction

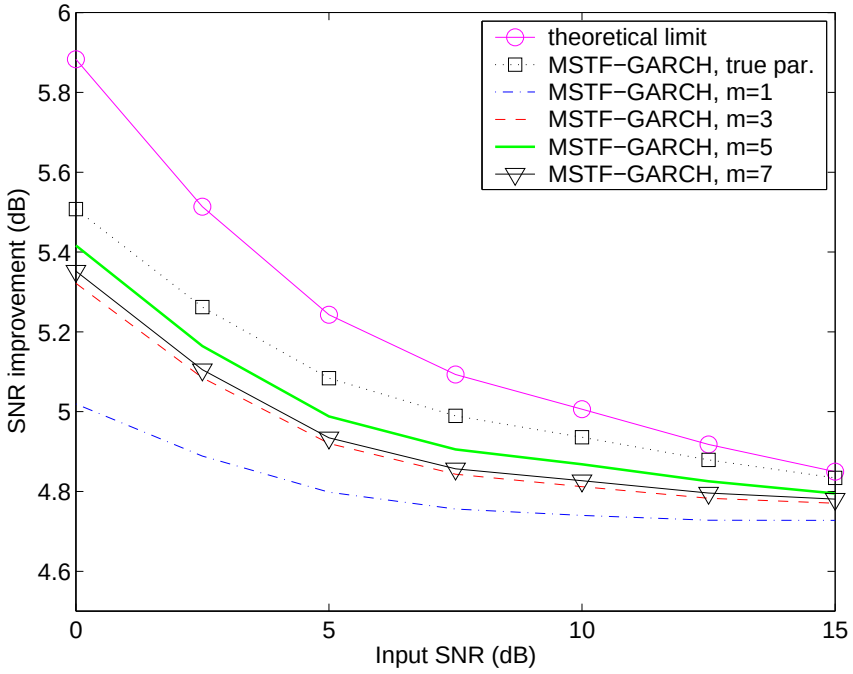
Experimental Results

Signal estimation - synthetic signals

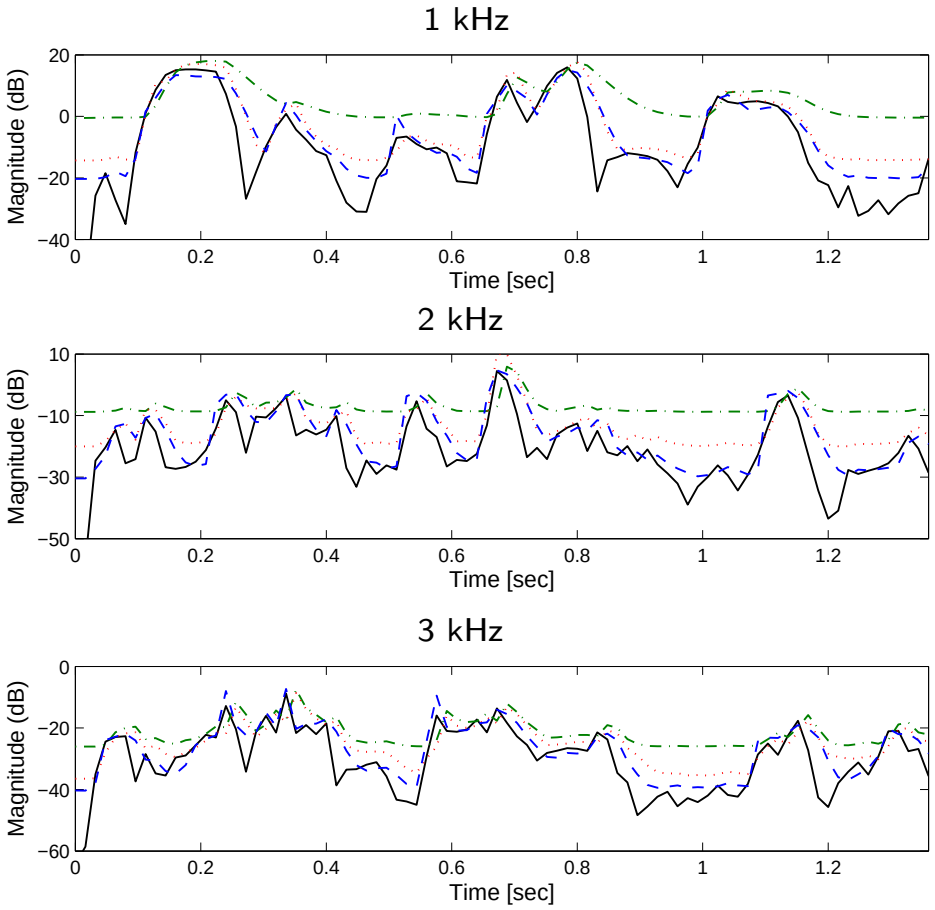
3-state model



5-state model



Estimated conditional variances at various frequencies:

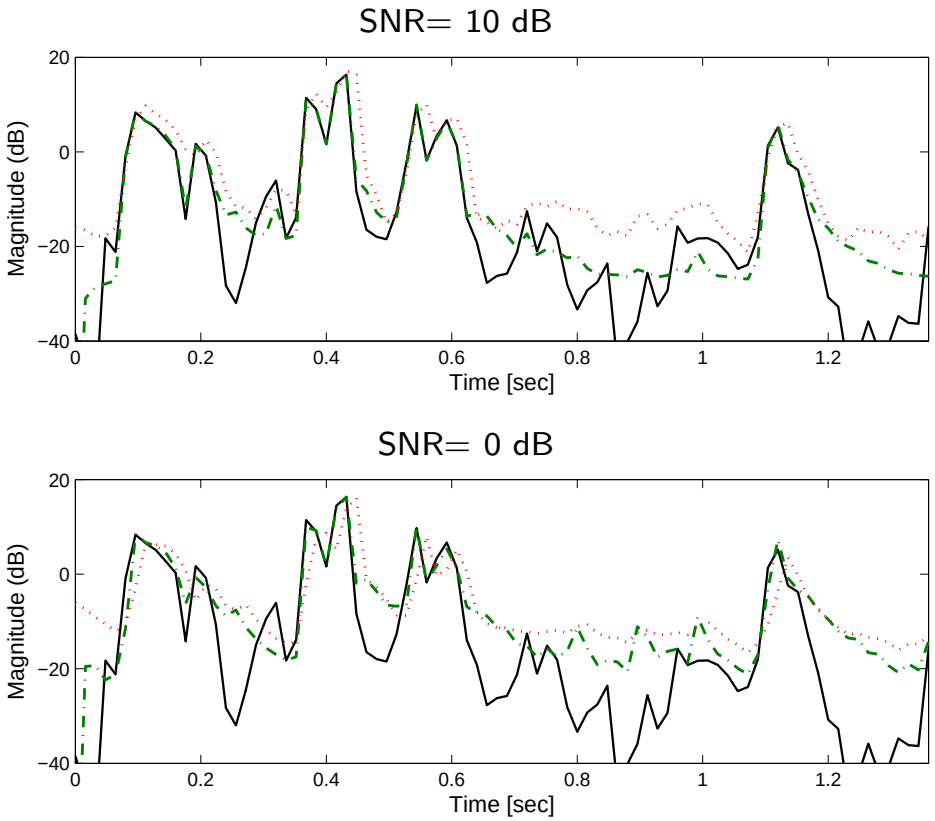


Solid line: Signal's squared absolute value
Dashed-dotted line: single-state model
Dotted line: 3-state model

Dashed line: 5-state model

Variance estimation (cont.)

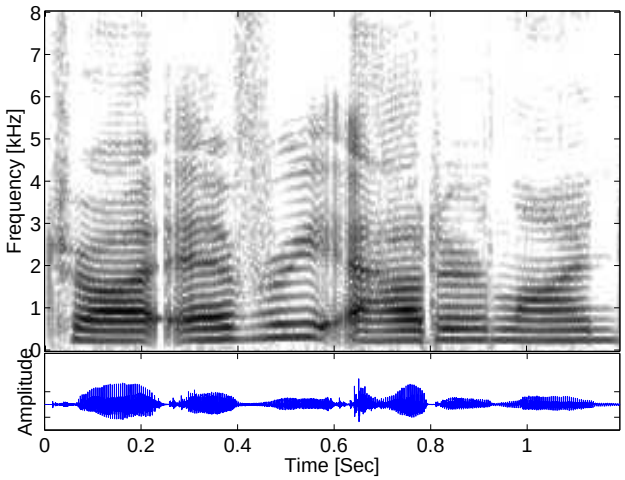
Estimated squared absolute values at frequency of 2 kHz:



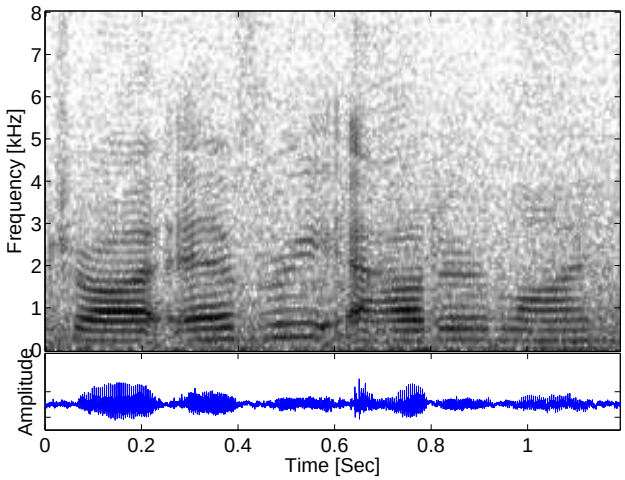
Solid line:
Speech signal's squared absolute value.
Dashed-dotted line:
5-state MSTF-GARCH model.
Dotted line:
Decision-directed approach.

Speech Enhancement - MS-GARCH Modeling

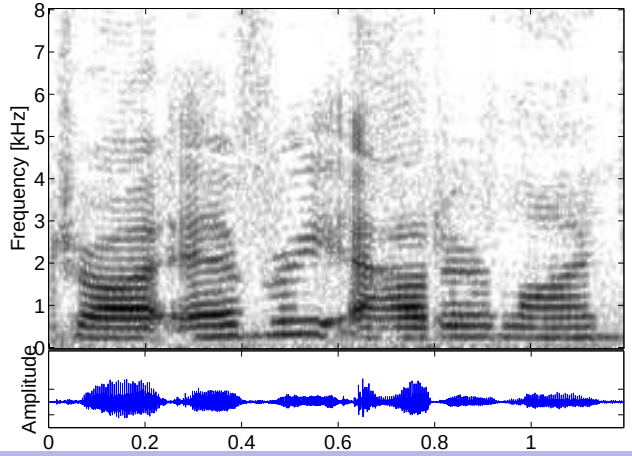
Clean signal



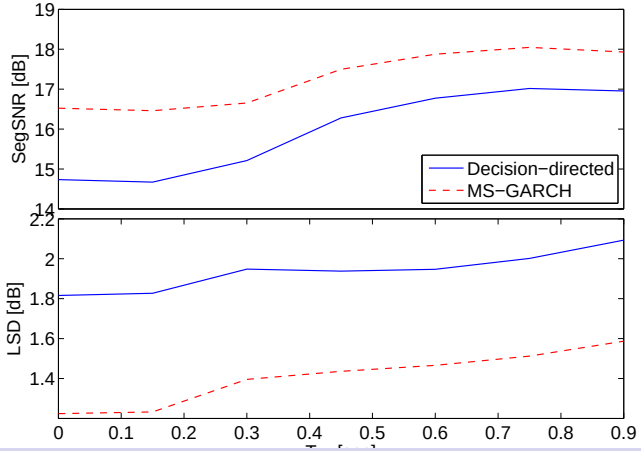
Speech corrupted by a factory noise with 5 dB SNR
(LSD= 6.68 dB, SegSNR= 0.05 dB)



Speech reconstructed by using a 4-state model
(LSD= 3.14 dB, SegSNR= 6.76 dB)

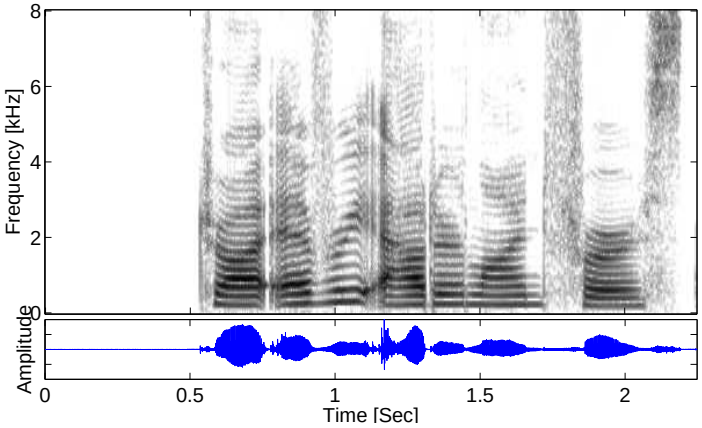


Comparison with DD
Reverberated speech with 20 dB SNR

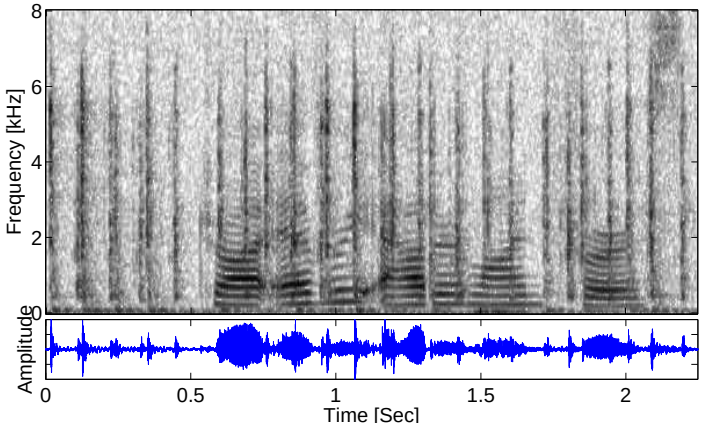


Keyboard-Typing Noise, Detection vs. Estimation

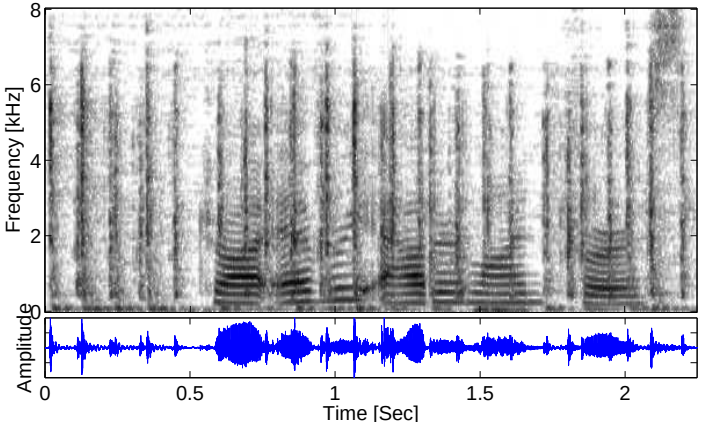
Clean signal



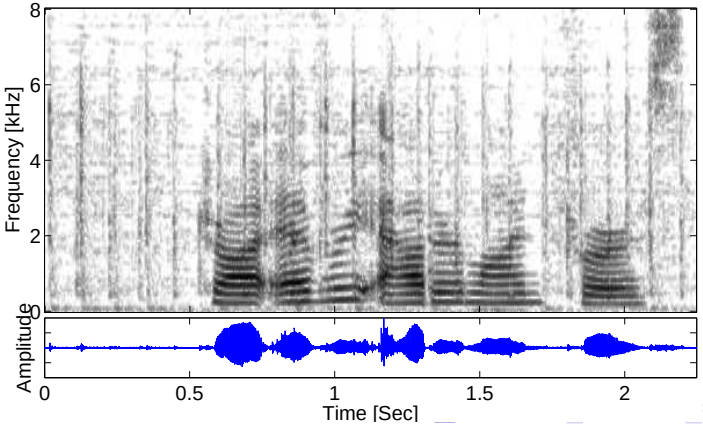
Noisy signal with 0.8 dB SNR
(LSD=7.69 dB, SegSNR=-2.23 dB)



OM-LSA
(LSD=6.77 dB, SegSNR=-1.31 dB)

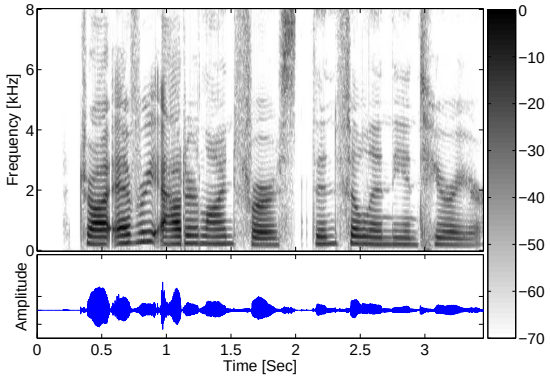


Detection and estimation
(LSD=2.52 dB, SegSNR=4.41 dB)

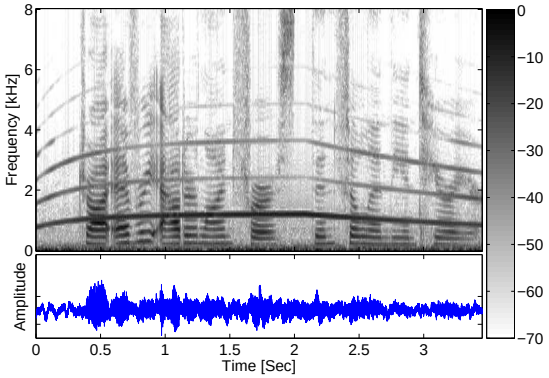


Car Environment with a Siren, Detection vs. Estimation

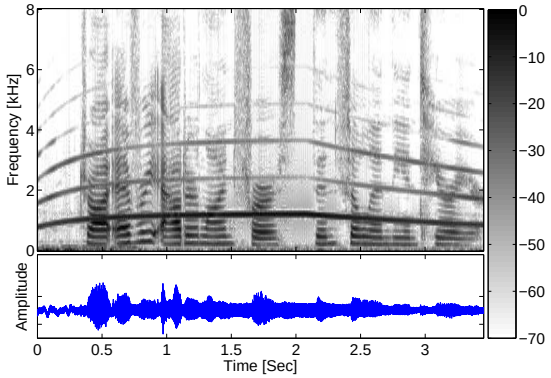
Clean signal



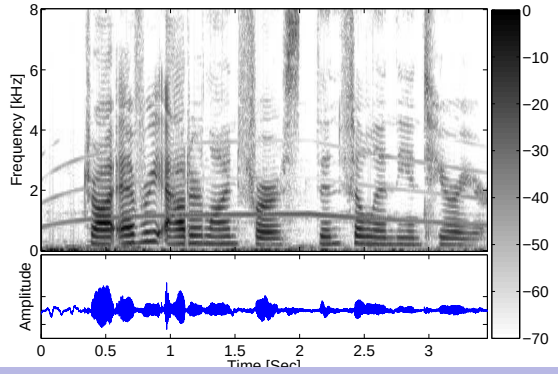
Noisy signal with -3 dB SNR
(LSD=6.56 dB, SegSNR=-5.95 dB)



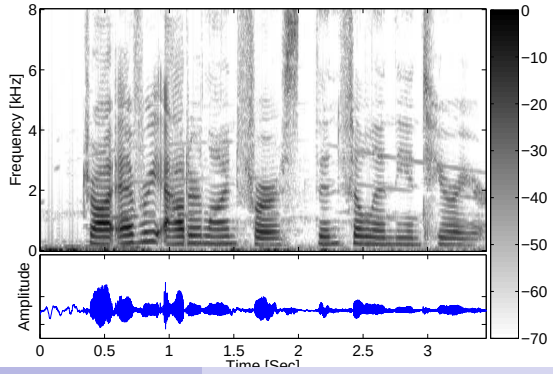
OM-LSA
(LSD=4.609 dB, SegSNR=-0.81 dB)

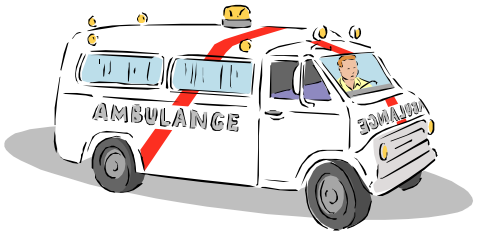


STSA with siren noise detector
(LSD=4.37 dB, SegSNR=-0.515 dB)

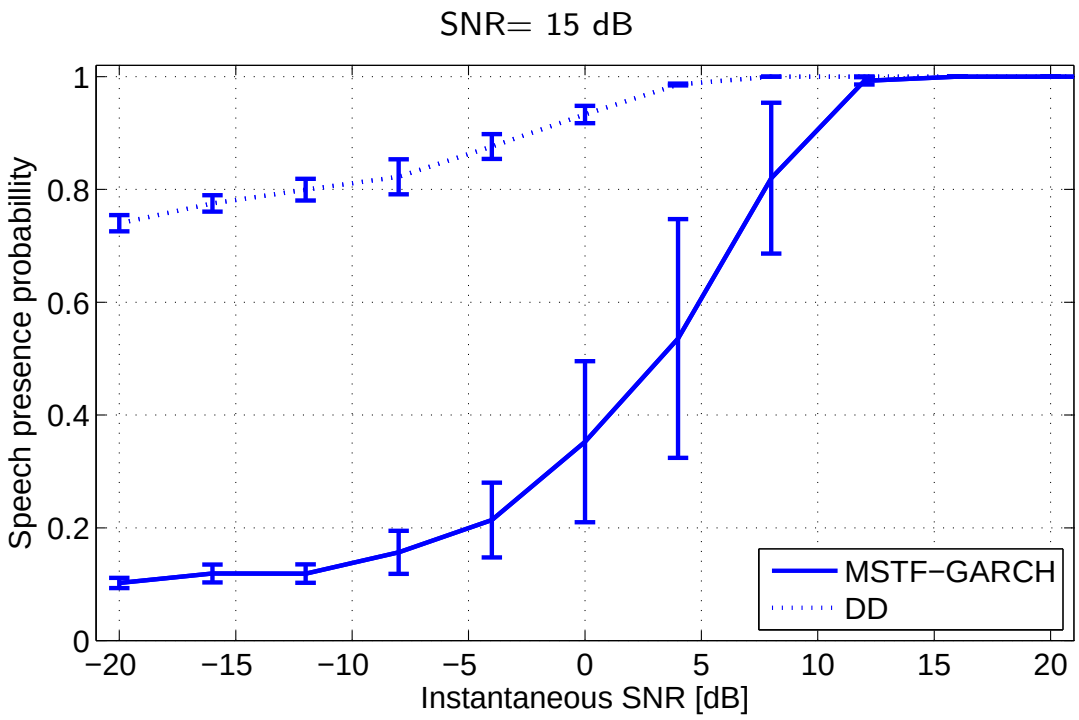


Sim. detection and estimation
(LSD=3.14 dB, SegSNR=6.50 dB)





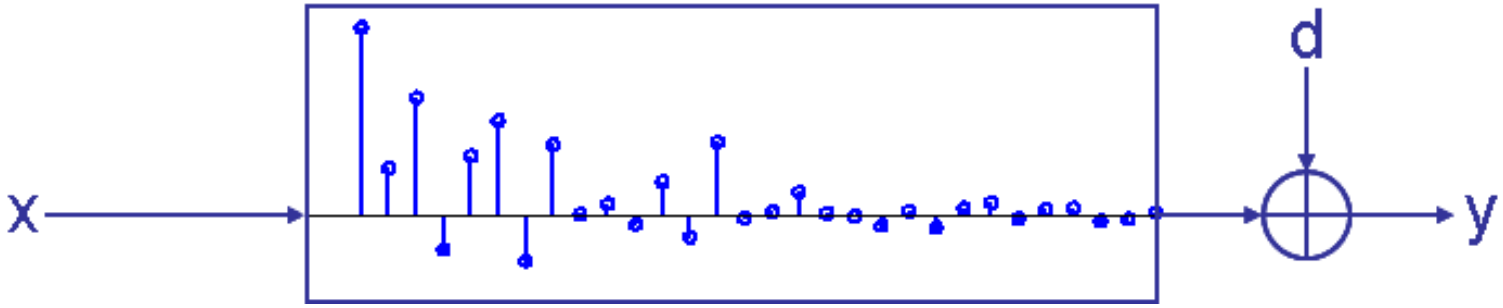
Voice Activity Detection Using MS-GARCH Model



- Higher dynamic range for the speech presence probabilities.
- Much lower probabilities for low energy coefficients.

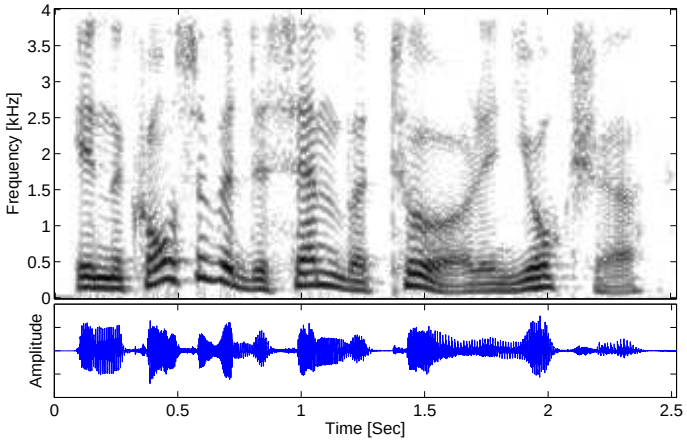
Speech Dereverberation

Noisy and reverberant speech:

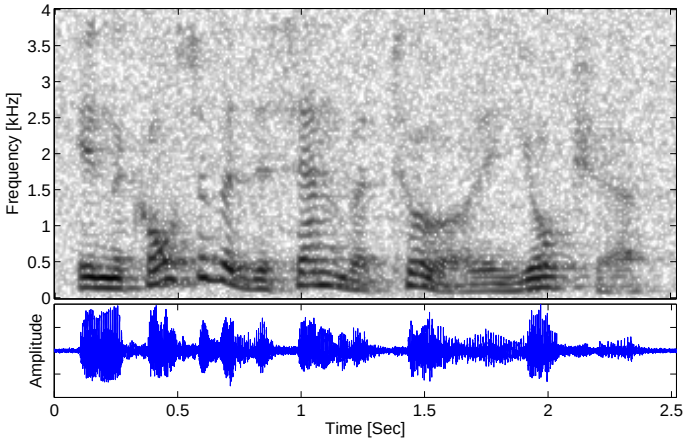


Speech Dereverberation

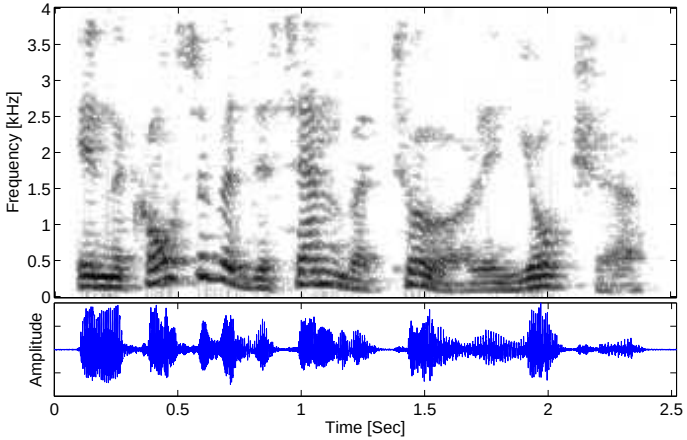
Clean signal



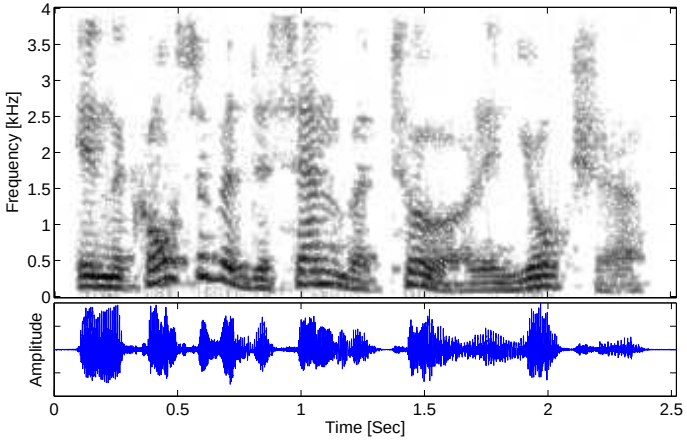
Corrupted signal
LSD=4.87 dB SegSNR=5.85 dB



DD-based algorithm
LSD=1.99 dB SegSNR=8.35 dB



Proposed algorithm
LSD=1.70 dB SegSNR=9.01 dB

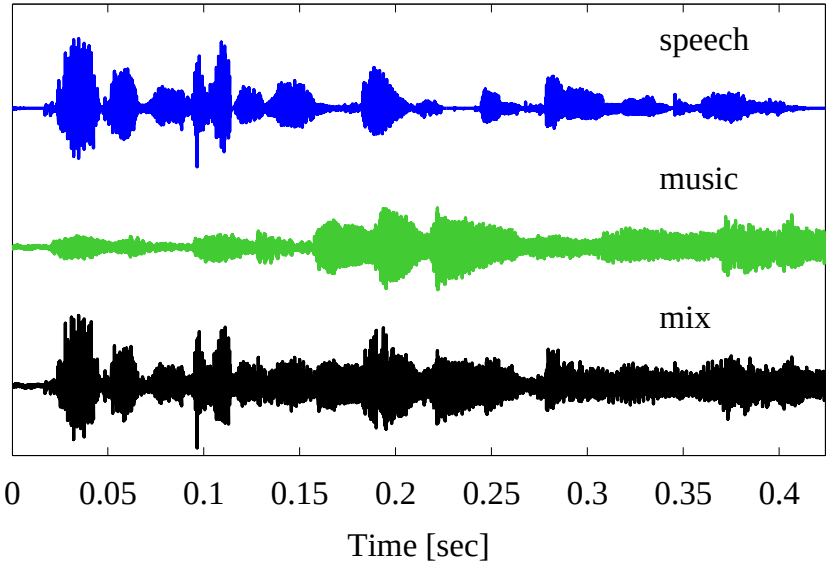


* Joint work with Habets and Gannot.

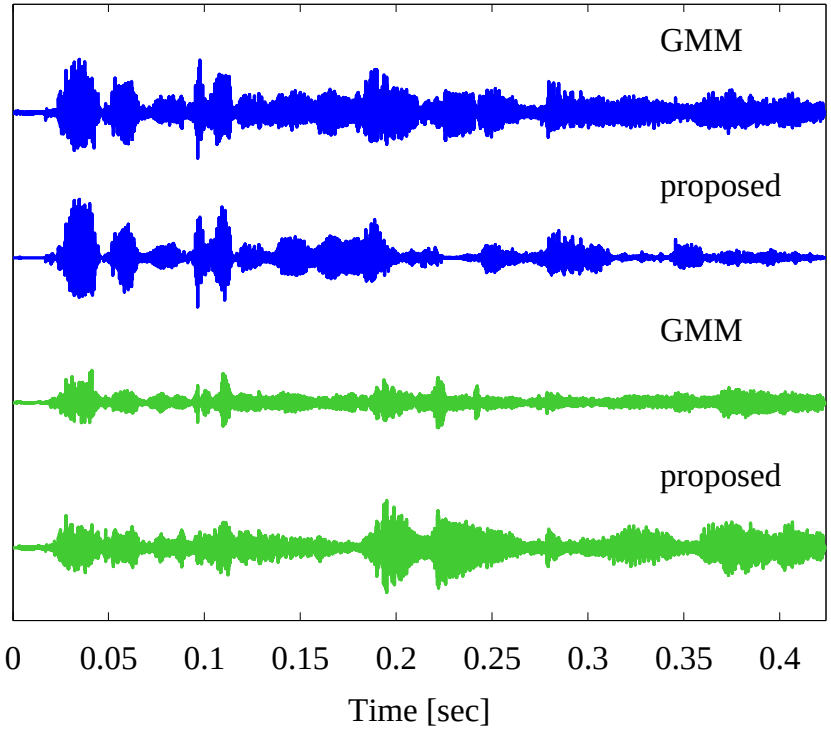
Single-Channel Blind Source Separation

GARCH modeling with Simultaneous Classification and Estimation:

Clean and mixed signals



Estimated signals



Summary

- A new formulation for the problem of speech enhancement based on a simultaneous detection and estimation approach.
- A novel statistical model for speech signals in the STFT domain.
- The statistical model and the detection and estimation approach may be applied to speech enhancement in highly nonstationary noise environments, speech dereverberation and single-sensor blind source separation.

Future Research:

- Improve multichannel algorithms using GARCH modeling and detection and estimation approach.
- Multivariate GARCH model with cross-band correlation.
- Optimizing the cost function and the trade-off parameters in case of multi-hypotheses.

Thank you!