Recursive Supervised Estimation of a Markov-Switching GARCH Process in the Short-Time Fourier Transform Domain

Ari Abramson, *Student Member, IEEE*, and Israel Cohen, *Senior Member, IEEE*

***Abstract*—In this paper, we introduce a Markov-switching generalized autoregressive conditional heteroscedasticity (GARCH) model for nonstationary processes with time-varying volatility structure in the short-time Fourier transform (STFT) domain. The expansion coefficients in the STFT domain are modeled as a multivariate complex GARCH process with Markov-switching regimes. The GARCH formulation parameterizes the correlation between sequential conditional variances while the Markov chain allows the process to switch between regimes of different GARCH formulations. We obtain a necessary and sufficient condition for the asymptotic wide-sense stationarity of the model, and develop a recursive algorithm for signal restoration in a noisy environment. The conditional variance is estimated by iterating propagation and update steps with regime conditional probabilities, while the model parameters are evaluated *a priori* from a training data set. Experimental results demonstrate the performance of the proposed algorithm.**

***Index Terms*—Generalized autoregressive conditional heteroscedasticity (GARCH), hidden Markov model, recursive estimation.**

I. Introduction

The generalized autoregressive conditional heteroscedasticity (GARCH) model is widely used in the field of econometrics for volatility forecast derivation of economic rates. This model, first introduced by Bollerslev [1] as a generalization of the ARCH model [2], explicitly parameterizes the time-varying volatility by using both recent conditional variances and recent squared innovations. GARCH models preserve the persistence of the process volatility in the sense that small variations tend to follow small variations and large variations tend to follow large variations. Incorporating GARCH models with hidden Markov chains, where each state (regime) of the chain implies a different GARCH behavior, extends the dynamic formulation of the model and enables a better fit for a process with a more complex time-varying volatility structure [3]–[5]. However, a major drawback of such models is that estimating the volatility with switching-regimes requires knowledge of the entire history of the process, including the regime path. Consequently, Cai [6] and Hamilton and Susmel [7] proposed a Markov-switching ARCH model, which avoids problems of path dependency in a noiseless environment. The conditional variance in ARCH models depends on previous observations only, so the Markov chain does not have to be known for constructing the conditional variance for a given regime. Gray [8] introduced a variant of Markov-switching GARCH model relying on the assumption that the conditional variance given current regime is dependent on the *expectation* of the previous conditional variances rather than their values. Accordingly, the conditional variance depends on some finite, state dependent, expected conditional variances via their conditional state probabilities. Klaassen [3] proposed modifying Gray's model by manipulating the current regime and all available observations while evaluating the expectation of previous conditional variances. A different method for reducing the dependency of the conditional variance on past regimes has recently been proposed by Haas, Mittnik, and Paolella [4]. Accordingly, a Markov chain governs the ARCH parameters while the autoregressive behavior of the conditional variance is subject to the assumption that past conditional variances are in the same regime as that of the current conditional variance. Gray, Klaassen and Haas *et al.* developed their variants of Markov-switching GARCH models for improved volatility forecasts of financial time-series under possible existence of shocks. They assumed that a process is observed in a noiseless environment so that its past observations provide a complete specification of its current conditional variance, for any given regime.

Recently, GARCH models have been employed for modeling speech signals in the time-frequency domain [9]–[11]. Speech signals in the short-time Fourier transform (STFT) domain demonstrate both "variability clustering" and heavy tail behavior similarly to financial time-series [11]. Motivated by these characteristics, it was proposed to model the conditional variance of speech signals in the STFT domain by a complex, K-dimensional GARCH model, with statistically independent elements (given past information) sharing the same GARCH specification. This time-frequency GARCH (TF-GARCH) model has been shown useful for speech enhancement applications, but it relies on the assumption that the model parameters are time-invariant. In [12], a GARCH model has been utilized in the time domain for speech recognition applications. The model parameters, characterizing the speech phonemes, are assumed speaker independent and time-varying. It was shown

Manuscript received September 28, 2005; revised October 15, 2006. This work was supported by the Israel Science Foundation by Grant 1085/05 . The associate editor coordinating the review of this manuscript and approving it for publication was Dr. P. Abry.

The authors are with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: aari@tx.technion.ac.il; icohen@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.894422

that estimating the GARCH specifications for each speech segment and using the parameters as part of the signal characteristics, speech recognition performance can be improved.

In this paper, we introduce a Markov-switching time-frequency GARCH (MSTF-GARCH) model which exploits the advantages of both the conditional heteroscedasticity structure of GARCH models and the time-varying characteristics of hidden Markov chains. Modeling probability density functions of speech signals by utilizing hidden Markov models has been found useful in speech recognition applications [13]–[15], and modeling the speech spectral coefficients as hidden Markov processes with a probability density prototype in each frame was applied to the problem of speech enhancement [16], [17]. Here we model the expansion coefficients of nonstationary random signals in the time-frequency domain as multivariate complex GARCH processes with Markov-switching regimes, and obtain a necessary and sufficient condition for the asymptotic wide-sense stationarity of the model. A corresponding recursive algorithm is developed for signal restoration in a noisy environment. The conditional variance is estimated by iterating propagation and update steps with regime conditional probabilities. The model parameters are estimated from a training data set prior to the signal restoration using maximum-likelihood (ML) approach, and the number of states is assumed to be known. We show that the derivation in [18] of bounds on the mean-square error (mse) of a composite source signal estimation is applicable for obtaining an upper bound on the mse of a single step MSTF-GARCH estimation. Experimental results demonstrate the improved performance of the proposed algorithm for restoration of MSTF-GARCH process compared to using an estimator which assumes a stationary process and compared to using an estimator which assumes a smaller number of regimes than the process actually has. Furthermore, it is demonstrated that the squared absolute values of speech coefficients in the STFT domain are better evaluated by using the MSTF-GARCH model than by using the decision-directed approach.

This paper is organized as follows. In Section II, we introduce the Markov-switching time-frequency GARCH model and obtain a necessary and sufficient condition for its asymptotic wide-sense stationarity. In Section III, we address the problem of signal estimation from noisy observations. In Section IV, we derive an upper bound on a single estimation step mse. In Section V, we address the problem of model estimation. Finally, in Section VI we provide some experimental results which demonstrate restoration of MSTF-GARCH process from noisy observations, and estimation of conditional variances and squared absolute values in the STFT domain from noisy speech signals.

II. Markov-Switching Time-Frequency GARCH Model

In this section, we briefly review the TF-GARCH model [11], and introduce a new time-frequency GARCH model with Markov-switching regimes, which allows further flexibility in the formulation of the time variation of the conditional variance.

A. Time-Frequency GARCH Model

Let $\{X_{tk} | t = 0, \ldots, T-1, k = 0, \ldots, K-1\}$ be the coefficients of a time-frequency transformation of a discrete-time signal x (e.g., STFT coefficients), where t is the time frame index and k is the frequency-bin index. Let $\mathbf{X}_t \triangleq [X_{t,0}, \ldots, X_{t,K-1}]'$ be the vector of spectral coefficients at time frame t, let $\mathcal{X}^\tau = \mathcal{X}_0^\tau \triangleq \{\mathbf{X}_t | t = 0, \ldots, \tau\}$ represent the set of spectral coefficients up to time τ, and let $\lambda_{tk|\tau} \triangleq E\left\{|X_{tk}|^2 | \mathcal{X}^\tau\right\}$ denote the *conditional variance* of the spectral coefficient at time-frequency bin (t, k), given the clean spectral coefficients up to time τ. Let $\{\mathbf{V}_t\} \in \mathbb{C}^K$ be a complex Gaussian random process with $\mathbf{V}_t \sim \mathcal{CN}(0, I_K)$, where I_K is a K-by-K identity matrix. A K-dimensional time-frequency GARCH model of order (p, q), is defined as follows [11]:

$$X_{tk} = \sqrt{\lambda_{tk|t-1}} V_{tk}, \quad k = 0, \ldots, K-1 \tag{1}$$

$$\boldsymbol{\lambda}_{t|t-1} = \zeta \cdot \mathbf{1} + \sum_{i=1}^{q} \alpha_i \mathbf{X}_{t-i} \odot \mathbf{X}_{t-i}^* + \sum_{j=1}^{p} \beta_j \boldsymbol{\lambda}_{t-j|t-j-1} \tag{2}$$

where $\mathbf{1}$ denotes a vector of ones, $\odot$ denotes a term-by-term multiplication and * denotes complex conjugation. The conditional variance vector, $\boldsymbol{\lambda}_{t|t-1} = E\left\{\mathbf{X}_t \odot \mathbf{X}_t^* | \mathcal{X}^{t-1}\right\}$, referred to as the *one-frame-ahead conditional variance* [11], is a linear function of the coefficients' past squared values and conditional variances, where

$$\begin{aligned} &\zeta > 0, \quad \alpha_i \geq 0, \ i = 1, \ldots, q, \\ &\beta_j \geq 0, \quad j = 1, \ldots, p \end{aligned} \tag{3}$$

are sufficient constraints for the positivity of the conditional variance [1]. The time-frequency GARCH has been introduced in [9] for modeling speech signals in the STFT domain, but the parameters of the GARCH model are assumed time invariant. Extending this model such that the model parameters may vary with time introduces additional flexibility in the model formulation, which may result in better characterization of speech signals and improved restoration in noisy environments.

B. MSTF-GARCH Formulation

Let S_t denote the (unobserved) state at time t and let s_t be a realization of S_t, assuming S_t is a first-order Markov chain. Let $\mathcal{I}^t \triangleq \{\mathcal{X}^t, \mathcal{S}^t\}$ denote all available information up to time t, which contains the clean signal coefficients and the regimes path up to time t, $\mathcal{S}^t \triangleq \{s_0, \ldots, s_t\}$. Denote by $\lambda_{tk|t-1,s_t} \triangleq E\left\{|X_{tk}|^2 | \mathcal{I}^{t-1}, s_t\right\}$ the one-frame-ahead conditional variance of the spectral coefficient X_{tk} given the information up to time $t-1$ and the chain state s_t. We assume that the spectral coefficients X_{tk} are generated by an m-state Markov-switching time-frequency GARCH process of order (p, q), denoted by $X_{tk} \sim MSTF-GARCH(p, q)$, which follows:

$$X_{tk} = \sqrt{\lambda_{tk|t-1,s_t}} V_{tk}, \quad k = 0, \ldots, K-1 \tag{4}$$

and the one-frame-ahead conditional variance evolves as follows:

$$\boldsymbol{\lambda}_{t|t-1,s_t} = \zeta_{s_t} \mathbf{1} + \sum_{i=1}^{q} \alpha_{i,s_t} \mathbf{X}_{t-i} \odot \mathbf{X}_{t-i}^* + \sum_{j=1}^{p} \beta_{j,s_t} \boldsymbol{\lambda}_{t-j|t-j-1,s_{t-j}} \tag{5}$$

where

$$\zeta_s > 0, \quad \alpha_{i,s} \geq 0, \ \beta_{j,s} \geq 0$$
$$i = 1, \dots, q, \quad j = 1, \dots, p, \ \ s = 1, \dots, m \tag{6}$$

are sufficient constraints for the positivity of the one-frame-ahead conditional variance. It follows from (4) and (5) that the conditional density of the coefficients depends on past values (through previous conditional variances) and also on the regime-path up to the current time. As considered in previous works on TF-GARCH, we assume that the model parameters are frequency-invariant. This restriction can be easily relaxed for the case of frequency (or subband) dependent parameters, i.e., $\zeta_{k,s}$, $\alpha_{i,k,s}$, and $\beta_{i,k,s}$, but the complexity of the model estimation then grows rapidly (see Section V).

GARCH models provide a rich class of possible parametrization of conditional heteroscedasticity (i.e., time-varying volatility) and the hidden Markov chain allows these GARCH formulations to switch along time. Volatility persistence naturally arises in a single-regime GARCH model. However, the existence of a Markov chain with different GARCH parameters allows the process to switch between regimes of different volatility formulations and different levels of volatility.

C. Stationarity of an MSTF-GARCH Process

The conditional variance of a GARCH process, and in particular of a Markov-switching GARCH process, changes recursively over time. Consequently, asymptotic wide-sence stationarity is required to ensure a finite second-order moment [3], [4], [19]. Necessary and sufficient conditions for the asymptotic stationarity of two variants of GARCH models with Markov-switching regimes have been derived in [19]. Those models generalize the models of Klaassen [3] and Haas *et al.* [4], but they both differ from our MSTF-GARCH model, which is a multivariate, complex valued process that entails the regime path for the construction of the conditional variance from past observations. A necessary and sufficient condition for asymptotic wide-sense stationarity of an MSTF-GARCH process has been derived in [20]. For the completeness of this paper we briefly summarize these results.

Assuming a stationary Markov chain with stationary probabilities $\pi_s = p(S_t = s)$, the unconditional variance of the process can be calculated using (4) and (5):

$$\begin{aligned} E\{\mathbf{X}_t \odot \mathbf{X}_t^*\} &= \sum_{s_t} \pi_{s_t} E\{\mathbf{X}_t \odot \mathbf{X}_t^* | s_t\} \\ &= \sum_{s_t} \pi_{s_t} E\{\boldsymbol{\lambda}_{t|t-1,s_t}\} \end{aligned} \tag{7}$$

where

$$\begin{aligned} E\{\boldsymbol{\lambda}_{t|t-1,s_t}\} = \zeta_{s_t}\mathbf{1} &+ \sum_{i=1}^{q} \alpha_{i,s_t} E\{\mathbf{X}_{t-i} \odot \mathbf{X}_{t-i}^* | s_t\} \\ &+ \sum_{j=1}^{p} \beta_{j,s_t} E\{\boldsymbol{\lambda}_{t-j|t-j-1,S_{t-j}} | s_t\} \end{aligned} \tag{8}$$

and

$$\begin{aligned} E\{\boldsymbol{\lambda}_{t-i|t-i-1,S_{t-i}} | s_t\} = &\sum_{s_{t-i}} p(s_{t-i}|s_t) \\ &\times E\{\boldsymbol{\lambda}_{t-i|t-i-1,s_{t-i}}\}. \end{aligned} \tag{9}$$

Note that $E\{\boldsymbol{\lambda}_{t|t-1,s_t}\}$ denotes the expected value of the conditional variance under the regime $S_t = s_t$, but $E\{\boldsymbol{\lambda}_{t|t-1,S_t}|\cdot\}$ denotes a conditional expectation of the conditional variance at time t where the active regime at that time is unknown. Since no prior information is given, we have

$$E\{\mathbf{X}_{t-i} \odot \mathbf{X}_{t-i}^* | s_t\} = \sum_{s_{t-i}} p(s_{t-i}|s_t) E\{\boldsymbol{\lambda}_{t-i|t-i-1,s_{t-i}}\} \tag{10}$$

and consequently we obtain [20]

$$\begin{aligned} E\{\boldsymbol{\lambda}_{t|t-1,s_t}\} = \zeta_{s_t}\mathbf{1} &+ \sum_{i=1}^{r}\sum_{s_{t-i}} (\alpha_{i,s_t} + \beta_{i,s_t}) \\ &\times \frac{\pi_{s_{t-i}}}{\pi_{s_t}} \{A^i\}_{s_{t-i},s_t} E\{\boldsymbol{\lambda}_{t-i|t-i-1,s_{t-i}}\} \end{aligned} \tag{11}$$

where $r \triangleq \max\{p,q\}$, $\alpha_{i,s_t} \triangleq 0 \ \forall\, i > q$, $\beta_{i,s_t} \triangleq 0 \ \forall\, i > p$, and A is the transition probabilities matrix, i.e., $\{A\}_{ij} \triangleq a_{ij} = p(S_t = j | S_{t-1} = i)$. Define m-by-m matrices $\mathcal{K}_i$, $i = 1, \dots, r$ with elements

$$\{\mathcal{K}_i\}_{s,\tilde{s}} \triangleq (\alpha_{i,s} + \beta_{i,s}) \frac{\pi_{\tilde{s}}}{\pi_s} \{A^i\}_{\tilde{s},s}, \quad s, \tilde{s} = 1, \dots, m \tag{12}$$

and an mr-by-mr matrix as follows

$$\Psi \triangleq \begin{bmatrix} \mathcal{K}_1 & \mathcal{K}_2 & & \cdots & & \mathcal{K}_r \\ I_m & 0 & & \cdots & & 0 \\ 0 & I_m & & & & \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ & & & & & \\ 0 & \cdots & & 0 & I_m & 0 \end{bmatrix}. \tag{13}$$

Let $\rho(\cdot)$ denote the spectral radius of a matrix, i.e., its largest eigenvalue in modulus, and let Φ be an m-by-m square matrix built from the mr-by-mr matrix $(I - \Psi)^{-1}$ such that $\{\Phi\}_{ij} = \left\{(I - \Psi)^{-1}\right\}_{ij}$, $i, j = 1, \dots, m$. Then a necessary and sufficient condition for asymptotic wide-sense stationarity of an MSTF-GARCH process is $\rho(\Psi) < 1$, and the asymptotic covariance matrix of the process is then a diagonal matrix (see [20] for a detailed proof):

$$\lim_{t\to\infty} E\{\mathbf{X}_t \mathbf{X}_t^H\} = (\boldsymbol{\pi}\Phi\boldsymbol{\zeta}) I_K \tag{14}$$

where $\boldsymbol{\zeta} \triangleq [\zeta_1, \dots, \zeta_m]'$, $\boldsymbol{\pi}$ is the row vector of the stationary probabilities of the Markov chain, and $(\cdot)^H$ denotes the Hermitian transpose operation.

This stationarity condition is a necessary and sufficient condition for the existence of a finite second-order moment of the process. It implies that in some regimes (but not in all of them) the conditional variance may grow over time (i.e., $\sum_i \alpha_{i,s} + \sum_j \beta_{j,s} > 1$ for some states s) but still the unconditional variance can be finite [19], [20].

III. Restoration of Noisy MSTF-GARCH Process

In this section, we develop a recursive algorithm for the restoration of MSTF-GARCH processes observed in additive stationary noise.

A hidden Markov process is a discrete-time finite-state Markov chain observed through a memoryless invariant channel, where the chain state is assumed to be hidden but the transition probabilities between sequential states are assumed to be known. As a consequence of the memoryless channel, the conditional density of the observed signal at time t (say $\mathbf{X}_t$) given the chain states_t, depends only on the given state and not on previous observations, i.e., the conditional density of a hidden Markov process (HMP) realizes $p(\mathbf{X}_t|s_t,\mathbf{X}_{t-1},\mathbf{X}_{t-2},\ldots) = p(\mathbf{X}_t|s_t)$. Combining GARCH models with hidden Markov chains, where each state is assumed to have a different GARCH formulation, introduces further complexity when trying to forecast or estimate the process, since the conditional variance of the process evolves as a function of previous conditional variances, as implied from (5). Consequently, the conditional density depends on the entire history of the process, i.e., past values and active states. To avoid this problem, several variants of GARCH processes with Markov-switching regimes have been proposed, e.g., [3], [4], [8]. These models formulate differently the conditional variance at any regime as dependent on past signal observations only. However, these variants of Markov-switching GARCH models have been developed for the purpose of forecasting volatility of financial time-series, assuming that the process is observed in a noiseless environment, and that all past clean signal values are given.

We use an MSTF-GARCH (1,1) model, as defined in (4) and (5), to model complex, nonstationary random signals and we develop a recursive signal estimation algorithm for restoring the clean signal and its second-order moment, from noisy observations. The order (1,1) is chosen for computational simplicity since higher (p, q)-orders imply strong dependency of successive conditional variances. Therefore, $p = q = 1$ is generally assumed for the applications of Markov-switching GARCH modeling, e.g., [3], [4], [6]–[8]. Let $\{X_{tk}\}$ and $\{D_{tk}\}$ denote the spectral coefficients of signal and uncorrelated additive noise signal, respectively, and let $Y_{tk} = X_{tk} + D_{tk}$ represent the observed signal. Let $\mathbf{X}_t$ be a K-dimensional complex-valued stochastic process, which evolves as an m-state first-order MSTF-GARCH, i.e., $\mathbf{X}_t \sim \mathit{MSTF\text{-}GARCH}(1,1)$, and let $\mathbf{D}_t$ represent a K-dimensional complex Gaussian random noise, $\mathbf{D}_t \sim \mathcal{CN}(0, R^d)$, with known diagonal covariance matrix $R^d = \mathrm{diag}\{\boldsymbol{\sigma}^2\}$. We assume that all MSTF-GARCH model parameters are known, i.e., the initial regimes probability $\boldsymbol{\pi}^{(0)}$, the probability transitions matrix A, and the GARCH (1,1) parameters in each of the m regimes. Let $\phi \triangleq \{\boldsymbol{\pi}^{(0)}, A, \zeta_1,\ldots,\zeta_m,\alpha_1,\ldots,\alpha_m,\beta_1,\ldots,\beta_m\}$ be the set of parameters which specifies the model, where for a first-order process we denote $\alpha_s \triangleq \alpha_{1,s}$ and $\beta_s \triangleq \beta_{1,s}$. In practice, the model parameters ϕ are estimated from a set of clean training signals as generally done with hidden Markov models [13], [16], [17], [21] while the covariance matrix of the noise process can be estimated using the minimum statistics [22] or the minima controlled recursive averaging algorithms [23], [24]. The problem of model estimation is addressed in Section V.

The spectral restoration problem is generally formulated as deriving an estimator $\hat{X}_{tk}$ for the spectral coefficients, such that the expected value of a certain distortion measure is minimized. We develop a recursive estimator for the signal's spectral coefficients and for their absolute squared values in the sense of minimum mean-square error (mmse), and we then extend this framework to signal restoration in the sense of mmse of the log-spectral amplitude (LSA), which is often used in speech enhancement applications, see for instance [23], [25].

Let $\mathcal{Y}^\tau = \mathcal{Y}_0^\tau \triangleq \{\mathbf{Y}_t | t = 0,\ldots,\tau\}$ be the set of observations up to time τ. The causal mmse estimator of the coefficients $\mathbf{X}_t$ given the noisy observations up to time t is obtained as follows:

$$E\left\{\mathbf{X}_t|\mathcal{Y}^t\right\} = \sum_{s_t} p\left(s_t|\mathcal{Y}^t\right) E\left\{\mathbf{X}_t|s_t,\mathcal{Y}^t\right\}. \tag{15}$$

Denote the state dependent, one-frame-ahead conditional covariance matrix of the clean signal as

$$R^x_{s_t} \triangleq E\left\{\mathbf{X}_t\mathbf{X}_t^H|s_t,\mathcal{I}^{t-1}\right\}. \tag{16}$$

Following the model formulation this covariance matrix is a function of $R^x_{s_{t-1}}$ and $\mathbf{X}_{t-1}$ only. However, the clean signal values are usually unavailable, nor the sequence of active states, so the evaluation of (16) requires the whole available observations. To overcome this problem, we assume that given current regime, past estimated conditional covariances are sufficient statistics for the conditional variance estimation [10]. Accordingly, given the set of estimated one-frame-ahead conditional variances $\hat{\Lambda}_t \triangleq \left\{\hat{\boldsymbol{\lambda}}_{t|t-1,S_t}|S_t = 1,\ldots,m\right\}$ which manipulates the observations up to time $t-1$, we may use the following signal estimator:

$$\hat{\mathbf{X}}_t = \sum_{s_t} p\left(s_t|\hat{\Lambda}_t,\mathbf{Y}_t\right) E\left\{\mathbf{X}_t|s_t,\hat{R}^x_{s_t},\mathbf{Y}_t\right\} \tag{17}$$

where under a Gaussian model

$$E\left\{\mathbf{X}_t|s_t,\hat{R}^x_{s_t},\mathbf{Y}_t\right\} = \hat{R}^x_{s_t}\left(\hat{R}^x_{s_t} + R^d\right)^{-1}\mathbf{Y}_t. \tag{18}$$

Note that $\hat{R}^x_{s_t}$ is a K-by-K diagonal matrix (since $\{V_{tk}\}$ are statistically independent) with the estimated state-dependent conditional variance $\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}$ on its diagonal. This state-dependent conditional variance can be recursively estimated in the mmse sense by calculating its conditional expectation under s_t given the observation $\mathbf{Y}_{t-1}$ and the previous set of estimated conditional variances:

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_{t|t-1,s_t} &\triangleq E\left\{\boldsymbol{\lambda}_{t|t-1,S_t}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right\}\\ &= \zeta_{s_t}\mathbf{1} + \alpha_{s_t}E\left\{\mathbf{X}_{t-1}\odot\mathbf{X}^*_{t-1}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right\}\\ &\quad + \beta_{s_t}E\left\{\boldsymbol{\lambda}_{t-1|t-2,S_{t-1}}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right\}.\end{aligned} \tag{19}$$

The conditional second-order moment in (19), can be obtained by

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_{t-1|t-1,s_t} &\triangleq E\left\{\mathbf{X}_{t-1}\odot\mathbf{X}^*_{t-1}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right\}\\ &= \sum_{s_{t-1}} p\left(s_{t-1}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right)\\ &\quad\times E\left\{\mathbf{X}_{t-1}\odot\mathbf{X}^*_{t-1}|s_{t-1},s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right\}\\ &= \sum_{s_{t-1}} p\left(s_{t-1}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right)\\ &\quad\times E\left\{\mathbf{X}_{t-1}\odot\mathbf{X}^*_{t-1}|s_{t-1},\hat{\boldsymbol{\lambda}}_{t-1|t-2,s_{t-1}},\mathbf{Y}_{t-1};\phi\right\}\\ &\triangleq \sum_{s_{t-1}} p\left(s_{t-1}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right)\hat{\boldsymbol{\lambda}}_{t-1|t-1,s_{t-1}}.\end{aligned} \tag{20}$$

The expected one-frame-ahead conditional variance in (19), given the one-frame-ahead regime, can be obtained by

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_{t-1|t-2,s_t} &\triangleq E\left\{\boldsymbol{\lambda}_{t-1|t-2,S_{t-1}}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right\}\\ &=\sum_{s_{t-1}} p\left(s_{t-1}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right)\\ &\quad\times E\left\{\boldsymbol{\lambda}_{t-1|t-2,S_{t-1}}|s_{t-1},s_t,\hat{\Lambda}_{t-1};\phi\right\}\\ &=\sum_{s_{t-1}} p\left(s_{t-1}|s_t,\hat{\Lambda}_{t-1},\mathbf{Y}_{t-1};\phi\right)\hat{\boldsymbol{\lambda}}_{t-1|t-2,s_{t-1}}.\end{aligned} \tag{21}$$

The third lines in (20) and in (21) rely on the fact that given all observations up to time $t-1$ and given the state s_{t-1}, the second-order moment of the process at that time, and also its conditional variance, are independent of any future state. Moreover, notice that $\hat{\boldsymbol{\lambda}}_{t|t,s_t}$ and $\hat{\boldsymbol{\lambda}}_{t|t,s_{t+1}}$ in (20) represent the expected second-order moment of the process based on information up to time t, given the chain state at the same time, and given the next state, respectively. Similarly, $\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}$ and $\hat{\boldsymbol{\lambda}}_{t|t-1,s_{t+1}}$ in (21) represent the expectation of the one-frame-ahead conditional variance at time t given the chain state s_t, and given the chain state at the next time step, respectively.

The mmse estimation of the process' second-order moment $\hat{\boldsymbol{\lambda}}_{t|t,s_t}$ in (20) given the estimated one-frame-ahead conditional variance of the same regime $\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}$ (21), can be obtained by

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_{t|t,s_t} &= E\left\{\mathbf{X}_t\odot\mathbf{X}_t^*|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t},\mathbf{Y}_t\right\}\\ &=\hat{R}^x_{s_t}\left(\hat{R}^x_{s_t}+R^d\right)^{-1}\\ &\quad\times\left[\boldsymbol{\sigma}^2+\hat{R}^x_{s_t}\left(\hat{R}^x_{s_t}+R^d\right)^{-1}(\mathbf{Y}_t\odot\mathbf{Y}_t^*)\right]\end{aligned} \tag{22}$$

for $s_t=1,\ldots,m$, similarly to the method in [10] applied to the case of a single-regime spectral GARCH. Following the notation in [10] we call (22) the *update* step as it updates the estimation of the signal's second-order moment at time t from its estimated one-frame-ahead conditional variance, using the new observation $\mathbf{Y}_t$. Substituting (20), (21) and (22) into (19) we obtain the *propagation* step which propagates ahead in time to obtain a conditional variance estimation at the next time, $t+1$ (assuming regime s_{t+1}), using the available information up to the current time t

$$\hat{\boldsymbol{\lambda}}_{t+1|t,s_{t+1}}=\zeta_{s_{t+1}}\mathbf{1}+\alpha_{s_{t+1}}\hat{\boldsymbol{\lambda}}_{t|t,s_{t+1}}+\beta_{s_{t+1}}\hat{\boldsymbol{\lambda}}_{t|t-1,s_{t+1}}. \tag{23}$$

for $s_{t+1}=1,\ldots,m$.

Let $\hat{\Lambda}^t\triangleq\left\{\hat{\Lambda}_0,\hat{\Lambda}_1,\ldots,\hat{\Lambda}_t\right\}$ be the set of the recursively estimated conditional variances up to time t, then we can manipulate all previous estimations to recursively evaluate the probability $p\left(s_{t-1}|s_t,\hat{\Lambda}^{t-1},\mathbf{Y}_{t-1};\phi\right)$ in (20) and (21) by

$$p\left(s_{t-1}|s_t,\hat{\Lambda}^{t-1},\mathbf{Y}_{t-1};\phi\right)=\frac{p\left(s_{t-1}|\hat{\Lambda}^{t-1},\mathbf{Y}_{t-1};\phi\right)a_{s_{t-1},s_t}}{p\left(s_t|\hat{\Lambda}^{t-1},\mathbf{Y}_{t-1};\phi\right)} \tag{24}$$

where

$$p\left(s_t|\hat{\Lambda}^{t-1},\mathbf{Y}_{t-1};\phi\right)=\sum_{s_{t-1}}p\left(s_{t-1}|\hat{\Lambda}^{t-1},\mathbf{Y}_{t-1};\phi\right)a_{s_{t-1},s_t}. \tag{25}$$

The conditional state probability at the right of (25) can be obtained by

$$\begin{aligned}p\left(s_t|\hat{\Lambda}^t,\mathbf{Y}_t;\phi\right) &=\frac{b\left(\mathbf{Y}_t,s_t|\hat{\Lambda}^t;\phi\right)}{b\left(\mathbf{Y}_t|\hat{\Lambda}^t;\phi\right)}\\ &=\frac{b\left(\mathbf{Y}_t|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)p\left(s_t|\hat{\Lambda}^{t-1},\mathbf{Y}_{t-1};\phi\right)}{\sum_{s_t}b\left(\mathbf{Y}_t|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)p\left(s_t|\hat{\Lambda}^{t-1},\mathbf{Y}_{t-1};\phi\right)}\end{aligned} \tag{26}$$

where $b(\cdot|\cdot)$ denotes a conditional density function. Specifically, $b\left(\mathbf{Y}_t|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)$ is the observation conditional density which is a complex normal distribution with zero-mean and $\hat{R}^x_{s_t}+R^d$ covariance matrix

$$\begin{aligned}b\left(\mathbf{Y}_t|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right) &=\frac{1}{\pi^K\left|\hat{R}^x_{s_t}+R^d\right|}\\ &\quad\times\exp\left\{\mathbf{Y}_t^H\left(\hat{R}^x_{s_t}+R^d\right)^{-1}\mathbf{Y}_t\right\}.\end{aligned} \tag{27}$$

Computing the conditional density $b\left(\mathbf{Y}_t|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)$ tends to be numerically unstable for large values of K since the diagonal values of its covariance matrix (i.e., $\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}$) are typically of the same order of magnitude. Therefore, $b\left(\mathbf{Y}_t|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)$ tends to zero or infinity exponentially fast as K increases. It is therefore useful to recursively evaluate a normalized density $\tilde{b}\left(\mathbf{Y}_t|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)$ as follows: see (28) at the bottom of the page, for $\tilde{k}=0,\ldots,K-1$ and substitute it into (26). As can be seen from (26), this normalization of $b\left(\mathbf{Y}_t|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)$ does not affect the value of $p\left(s_t|\hat{\Lambda}^t,\mathbf{Y}_t;\phi\right)$.

The causal one-frame-ahead conditional variance and the conditional second-order moment of the process can be

$$\tilde{b}\left(Y_{t,0},\ldots,Y_{t\tilde{k}}|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)=\frac{\tilde{b}\left(Y_{t,0},\ldots,Y_{t,\tilde{k}-1}|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)b\left(Y_{t,\tilde{k}}|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)}{\sum_{s_t}\tilde{b}\left(Y_{t,0},\ldots,Y_{t,\tilde{k}-1}|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)b\left(Y_{t,\tilde{k}}|s_t,\hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)} \tag{28}$$

TABLE I
VECTOR FORM OF THE RECURSIVE MSTF-GARCH SIGNAL ESTIMATION

$$
\begin{aligned}
&\textit{Initialization}:\\
&\quad \boldsymbol{\rho}_{-1}(\mathbf{s}_0) = \boldsymbol{\pi}\\
&\quad \hat{\boldsymbol{\lambda}}_{-1,k|-2,\mathbf{s}_0} = \hat{\boldsymbol{\lambda}}_{-1,k|-1,\mathbf{s}_0} = \mathbf{0}_{m\times 1}, \quad k = 0,...,K-1\\
&\\
&\textit{for } t = 0,...,T-1\\
&\quad \hat{\boldsymbol{\lambda}}_{tk|t-1,\mathbf{s}_t} = \boldsymbol{\zeta} + \boldsymbol{\alpha} \odot \hat{\boldsymbol{\lambda}}_{t-1,k|t-1,\mathbf{s}_t} + \boldsymbol{\beta} \odot \hat{\boldsymbol{\lambda}}_{t-1,k|t-2,\mathbf{s}_t}, \quad k = 0,...,K-1\\
&\quad \hat{\boldsymbol{\lambda}}_{tk|t,\mathbf{s}_t} = \hat{\boldsymbol{\lambda}}_{tk|t-1,\mathbf{s}_t} \odot \left[\sigma_k^2 \mathbf{1} + (\hat{\boldsymbol{\lambda}}_{tk|t-1,\mathbf{s}_t} \cdot |Y_{tk}|^2)(\div)(\hat{\boldsymbol{\lambda}}_{tk|t-1,\mathbf{s}_t} + \sigma_k^2 \mathbf{1})\right]\\
&\qquad\qquad (\div)(\hat{\boldsymbol{\lambda}}_{tk|t-1,\mathbf{s}_t} + \sigma_k^2 \mathbf{1}), \quad k = 0,...,K-1\\
&\quad b(\mathbf{Y}_t|s_t, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t}) = \pi^{-K} |\hat{R}^x_{s_t} + R_d|^{-1} \exp\left\{-\mathbf{Y}_t^H (\hat{R}^x_{s_t} + R_d)^{-1} \mathbf{Y}_t\right\}, \quad s_t = 1,...,m\\
&\quad B_t \triangleq \mathrm{diag}\left\{\mathbf{b}(\mathbf{Y}_t|\mathbf{s}_t, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t})\right\}\\
&\quad \boldsymbol{\rho}_t(\mathbf{s}_t) = B_t\, \boldsymbol{\rho}_{t-1}(\mathbf{s}_t) \left[\mathbf{1}' B_t\, \boldsymbol{\rho}_{t-1}(\mathbf{s}_t)\right]^{-1}\\
&\quad \boldsymbol{\rho}_t(\mathbf{s}_{t+1}) = A' \boldsymbol{\rho}_t(\mathbf{s}_t)\\
&\quad \textit{for } i = 1,...,m : \mathbf{c}_t^{(i)} = \mathbf{a}^{(i)} \odot \boldsymbol{\rho}_t(\mathbf{s}_t) / \rho_t(s_{t+1} = i)\\
&\quad \hat{\boldsymbol{\lambda}}_{tk|t,\mathbf{s}_{t+1}} = C_t'\, \hat{\boldsymbol{\lambda}}_{tk|t,\mathbf{s}_t}, \quad k = 0,...,K-1\\
&\quad \hat{\boldsymbol{\lambda}}_{tk|t-1,\mathbf{s}_{t+1}} = C_t'\, \hat{\boldsymbol{\lambda}}_{tk|t-1,\mathbf{s}_t}, \quad k = 0,...,K-1\\
&\quad \hat{\mathbf{X}}_{t|t,s_t} = \hat{R}^x_{s_t} (\hat{R}^x_{s_t} + R^d)^{-1} \mathbf{Y}_t\\
&\quad \hat{X}_{tk} = \boldsymbol{\rho}_t'(\mathbf{s}_t) \hat{X}_{tk|t,s_t}, \quad k = 0,...,K-1
\end{aligned}
$$

obtained by

$$
\begin{aligned}
\hat{\boldsymbol{\lambda}}_{t|t-1} &= \sum_{s_t} p\left(s_t|\hat{\Lambda}^t, \mathbf{Y}_t\right) E\left\{\boldsymbol{\lambda}_{t|t-1,S_t}|s_t, \hat{\Lambda}_t\right\}\\
&= \sum_{s_t} p\left(s_t|\hat{\Lambda}^t, \mathbf{Y}_t\right) \hat{\boldsymbol{\lambda}}_{t|t-1,s_t} \qquad (29)\\
\hat{\boldsymbol{\lambda}}_{t|t} &= \sum_{s_t} p\left(s_t|\hat{\Lambda}^t, \mathbf{Y}_t\right) E\left\{\mathbf{X}_t \odot \mathbf{X}_t^* | s_t, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t}, \mathbf{Y}_t\right\}\\
&= \sum_{s_t} p\left(s_t|\hat{\Lambda}^t, \mathbf{Y}_t\right) \hat{\boldsymbol{\lambda}}_{t|t,s_t} \qquad (30)
\end{aligned}
$$

while a state smoothing (i.e., noncausal state probability estimation) for the path-dependent MSTF-GARCH model has been derived in [26] and may be employed for noncausal estimation.

The causal recursive mmse signal restoration algorithm, presented in (17) to (26), has a compact vector form with respect to the regimes vector. Let $\mathbf{s}_t \triangleq [S_t = 1, \ldots, S_t = m]'$ be the regimes vector at time t, let $\boldsymbol{\rho}_t(\mathbf{s}_\tau) \triangleq \left[p\left(S_\tau = 1|\hat{\Lambda}^t, \mathbf{Y}_t\right), \ldots, p\left(S_\tau = m|\hat{\Lambda}^t, \mathbf{Y}_t\right)\right]'$ be the probabilities of the regimes vector $\mathbf{s}_\tau$, conditioned on all observations up to frame t. Let C_t be a regimes probability matrix at time t conditioned on the next regime and all available observations up to time t, i.e., $c_{t,ij} = p\left(S_t = i|S_{t+1} = j, \hat{\Lambda}^t, \mathbf{Y}_t\right)$, $i,j = 1,\ldots,m$. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent the vectors of the m regimes' GARCH parameters, i.e., $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_m]'$ and $\boldsymbol{\beta} \triangleq [\beta_1, \ldots, \beta_m]'$. Let $\hat{\boldsymbol{\lambda}}_{tk|\tau_1,\mathbf{s}_{\tau_2}} \triangleq \left[\hat{\lambda}_{tk|\tau_1,S_{\tau_2}=1}, \ldots, \hat{\lambda}_{tk|\tau_1,S_{\tau_2}=m}\right]'$ be an $m \times 1$ vector of the kth index estimated conditional variances based on observations up to time τ_1, and the corresponding m regimes vector $\mathbf{s}_{\tau_2}$. Denote by $\mathbf{a}^{(i)}$ and $\mathbf{c}_t^{(i)}$ the ith column of matrices A and C_t, respectively, and let $(\div)$ denote a term-by-term division of two vectors. A step-by-step vector form of the causal signal estimation procedure is described in Table I.

The algorithm, summarized in Table I, estimates both the spectral coefficients and their conditional variance in the mmse sense. A more general signal enhancement problem is formulated as minimization of the following distortion measure:

$$E\left\{|f(X_{tk}) - f(\hat{X}_{tk})|^2 | \mathcal{Y}^t\right\} \qquad (31)$$

where $f(X)$ is a Borel integrable function. The estimator can be found from

$$f\left(\hat{X}_{tk}\right) = E\left\{f(X_{tk}) | \mathcal{Y}^t\right\} \qquad (32)$$

where

$$E\left\{f(X_{tk}) | \mathcal{Y}^t\right\} = \sum_{s_t} p\left(S_t = s_t | \mathcal{Y}^t\right) E\left\{f(X_{tk}) | s_t, \mathcal{Y}^t\right\}. \qquad (33)$$

The log-spectral amplitude mmse estimator, obtained by substituting $f(X) = \log|X|$ into (32), is of particular importance in speech enhancement applications, see for instance [23] and [25]. The LSA estimator [25] is given by

$$|\widehat{X_{tk}}| = \exp\left(E\left\{\log|X_{tk}| | \lambda_{tk}, Y_{tk}\right\}\right) = G(\xi_{tk}, \vartheta_{tk}) |Y_{tk}| \qquad (34)$$

where

$$\xi_{tk} \triangleq \frac{\lambda_{tk}}{\sigma_{tk}^2}, \quad \gamma_{tk} \triangleq \frac{|Y_{tk}|^2}{\sigma_{tk}^2}, \quad \vartheta_{tk} \triangleq \frac{\gamma_{tk}\xi_{tk}}{1+\xi_{tk}} \qquad (35)$$

and

$$G(\xi, \vartheta) = \frac{\xi}{1+\xi} \exp\left(\frac{1}{2} \int_\vartheta^\infty \frac{e^{-t}}{t} dt\right). \qquad (36)$$

ξ_{tk} and γ_{tk} represent the *a priori* and *a posteriori* SNRs, respectively, [27].

By substituting (34) into (33), and combining the result with the phase of the noisy signal [25], we obtain the spectral coefficient estimator in the mmse-LSA sense

$$\hat{X}_{tk} = Y_{tk} \prod_{s_t} G\left(\hat{\xi}_{tk,s_t}, \hat{\vartheta}_{tk,s_t}\right)^{p(s_t|\mathcal{Y}^t)} \tag{37}$$

where

$$\hat{\xi}_{tk,s_t} \triangleq \frac{\hat{\lambda}_{tk|t,s_t}}{\sigma^2_{tk}}, \quad \hat{\vartheta}_{tk,s_t} \triangleq \frac{\gamma_{tk}\hat{\xi}_{tk,s_t}}{1+\hat{\xi}_{tk,s_t}} \tag{38}$$

and $p(s_t|\mathcal{Y}^t)$ is recursively estimated using (26).

IV. Estimation Efficiency

In this section we analyze the mse of a one step ahead mmse estimation using the proposed recursive algorithm. The recursive formulation of the MSTF-GARCH yields an accumulated error in the estimation of the variance and the signal. However, for each regime and in each frame the algorithm evaluates the conditional variance as a weighted sum of previous estimated conditional variances and squared absolute values (20), (21). These weights are proportional to the conditional densities $b\left(\mathbf{Y}_t|s_t, \hat{\boldsymbol{\lambda}}_{t|t-1,s_t}\right)$ in (26). Consequently, an over estimation of the conditional variance on a specific frame can be followed in the algorithm by giving a high probability (i.e., higher weight) to a regime with small parameters which compensates the previous over estimation. Similarly, an under estimation of the conditional variance can be compensated by giving a high probability to a regime with large parameters.

Assume that the process is observed perfectly (without noise) up to time $t-1$ and that the regime path is known up to that time. Then, $\boldsymbol{\lambda}_{t-1|t-2,s_{t-1}}$ can be calculated by (5). Following Ephraim and Merhav [18] which derive bounds for the mse of a composite source signal estimation, we assume that: (i) the Markov chain is stationary and the necessary and sufficient condition for a bounded variance is satisfied; (ii) $\hat{\boldsymbol{\lambda}}_{t|t,s_t}$ is square integrable with respect to $b\left(\mathbf{Y}_t|\boldsymbol{\lambda}_{t|t-1,s_t}\right)$ and $b\left(\mathbf{Y}_t|\boldsymbol{\lambda}_{t|t-1,\tilde{s}_t}\right)$; and (iii) the regime transition probabilities are positive, i.e., $a_{ij} \geq a_{\min} > 0 \; \forall\, i,j = 1,\ldots,m$.

The one-step-ahead mmse estimator (17) is unbiased in the sense that $E\left\{\hat{\mathbf{X}}_t\right\} = E\left\{\mathbf{X}_t\right\}$, and following [18] we obtain an upper bound for the variance of the one-step-ahead estimation error, assuming that the process is observed with an additive, independent stationary noise. The one-step-ahead mse is given by

$$\overline{e_t^2} \triangleq \frac{1}{K}\mathrm{tr}E\left\{\left(\mathbf{X}_t - \hat{\mathbf{X}}_t\right)\left(\mathbf{X}_t - \hat{\mathbf{X}}_t\right)^H\right\} \tag{39}$$

where the signal estimator $\hat{\mathbf{X}}_t$ follows:

$$\begin{aligned}\hat{\mathbf{X}}_t &= E\left\{\mathbf{X}_t|\mathcal{I}^{t-1}, \mathbf{Y}_t\right\}\\ &= E\left\{\mathbf{X}_t|s_{t-1}, \boldsymbol{\lambda}_{t-1|t-2,s_{t-1}}, \mathbf{X}_{t-1}, \mathbf{Y}_t\right\}\\ &= E\left\{\mathbf{X}_t|\Lambda_t, \mathbf{Y}_t\right\}.\end{aligned} \tag{40}$$

Under the above assumptions, the mse can be written as [18, eqs. (13)–(17)]

$$\overline{e_t^2} = \overline{\mu_t^2} + \overline{\eta_t^2} \tag{41}$$

where

$$\begin{aligned}\overline{\mu_t^2} &\triangleq \frac{1}{K}\mathrm{tr}E\left\{\mathrm{cov}\left(\mathbf{X}_t|\Lambda_t, s_t, \mathbf{Y}_t\right)\right\}\\ &= \frac{1}{K}\mathrm{tr}E\left\{\mathrm{cov}\left(\mathbf{X}_t|\boldsymbol{\lambda}_{t|t-1,s_t}, s_t, \mathbf{Y}_t\right)\right\}\end{aligned} \tag{42}$$

$$\overline{\eta_t^2} \triangleq \frac{1}{2}\sum_{s_t\neq\tilde{s}_t} E\Big\{p\left(s_t|s_{t-1}, \Lambda_t, \mathbf{Y}_t\right) \times p\left(\tilde{s}_t|s_{t-1}, \Lambda_t, \mathbf{Y}_t\right) g\left(s_t, \tilde{s}_t, \Lambda_t, \mathbf{Y}_t\right)\Big\} \tag{43}$$

and

$$\begin{aligned}g\left(s_t, \tilde{s}_t, \Lambda_t, \mathbf{Y}_t\right) &\triangleq \frac{1}{K}\mathrm{tr}\Big\{\left(\hat{\mathbf{X}}_{t|t,s_t} - \hat{\mathbf{X}}_{t|t,\tilde{s}_t}\right) \times \left(\hat{\mathbf{X}}_{t|t,s_t} - \hat{\mathbf{X}}_{t|t,\tilde{s}_t}\right)^H\Big\}\\ &= g\left(s_t, \tilde{s}_t, \boldsymbol{\lambda}_{t|t-1,s_t}, \boldsymbol{\lambda}_{t|t-1,\tilde{s}_t}, \mathbf{Y}_t\right).\end{aligned} \tag{44}$$

The state probabilities in (43) can be evaluated using (26)

$$p\left(s_t|s_{t-1}, \Lambda_t, \mathbf{Y}_t\right) = \frac{b\left(\mathbf{Y}_t|s_t, \boldsymbol{\lambda}_{t|t-1,s_t}\right) a_{s_{t-1},s_t}}{\sum_{s_t} b\left(\mathbf{Y}_t|s_t, \boldsymbol{\lambda}_{t|t-1,s_t}\right) a_{s_{t-1},s_t}} \tag{45}$$

and the signal estimate given the state s_t is given by

$$\hat{\mathbf{X}}_{t|t,s_t} = W_{s_t}\mathbf{Y}_t \tag{46}$$

where W_{s_t} is the conditional Wiener filter: $W_{s_t} \triangleq R^x_{s_t}\left(R^x_{s_t} + R^d\right)^{-1}$. Substituting (46) into (44), we have

$$\begin{aligned}g\left(s_t, \tilde{s}_t, \Lambda_t, \mathbf{Y}_t\right) &= \frac{1}{K}\mathbf{Y}_t^H\left(W_{s_t} - W_{\tilde{s}_t}\right)^H\left(W_{s_t} - W_{\tilde{s}_t}\right)\mathbf{Y}_t\\ &\triangleq \frac{1}{K}\mathbf{Y}_t^H W^2_{\tilde{s}_t s_t}\mathbf{Y}_t.\end{aligned} \tag{47}$$

The one-step-ahead mse, $\overline{e_t^2}$, is decomposed into two positive terms, $\overline{\mu_t^2}$ and $\overline{\eta_t^2}$. The first is the mse of the estimator $\hat{\mathbf{X}}_{t|t,s_t}$ which relies on knowing the true regime at time t, and, therefore, it is the optimal estimator in the mmse sense. This term is evaluated by substituting (42) into (46)

$$\overline{\mu_t^2} = \frac{1}{K}\mathrm{tr}\sum_{s_t} a_{s_{t-1}s_t} W_{s_t} R^d. \tag{48}$$

The second term $\overline{\eta_t^2}$ is a weighted sum of cross error terms which depend on pairs of the process regimes. This term is difficult to evaluate, but it is upper bounded by [18, eqs. (18) and (23)]

$$\overline{\eta_t^2} \leq \frac{1}{2}\sum_{s_t\neq\tilde{s}_t} a_{\min}^{-2}\left(I_{s_t}(\tilde{s}_t) + I_{\tilde{s}_t}(s_t)\right) \tag{49}$$

where

$$\begin{aligned}I_{s_t}(\tilde{s}_t) \leq \frac{1}{K}\sum_{s_t\neq\tilde{s}_t}\mathrm{tr}\left\{W^2_{s_t\tilde{s}_t}Q_{\tilde{s}_t}\right\} &\times \Bigg(\left|R_\lambda(s_t,\tilde{s}_t)\right|\cdot\left|Q_{s_t}\right|^{-\lambda}\cdot\left|Q_{\tilde{s}_t}\right|^{\lambda-1}\\ &+ \frac{\mathrm{tr}\left\{W^2_{s_t\tilde{s}_t}R_\lambda(s_t,\tilde{s}_t)\right\}}{\mathrm{tr}\left\{W^2_{s_t\tilde{s}_t}Q_{\tilde{s}_t}\right\}}\Bigg), \quad \lambda > 0\end{aligned} \tag{50}$$

[18, eqs. (31)-(39) and (54)–(60)], Q_{s_t} denotes the covariance matrix of the noisy signal given the regime s_t, and $R_\lambda\left(s_t, \tilde{s}_t\right)$ is defined by

$$R_\lambda\left(s_t, \tilde{s}_t\right) \triangleq \left[\lambda Q_{s_t}^{-1} + (1-\lambda) Q_{\tilde{s}_t}^{-1}\right]^{-1}. \qquad (51)$$

In the derivation of (50) it is assumed that $R_\lambda\left(s_t, \tilde{s}_t\right)$ is positive definite [18]. Since Q_{s_t} is a diagonal matrix with positive eigenvalues, $R_\lambda\left(s_t, \tilde{s}_t\right)$ is positive definite for any $0 < \lambda < 1$. Substituting (50) into (49) and using the diagonality of the covariance matrices, we obtain an upper bound for the cross error term

$$\begin{aligned} \overline{\eta_t^2} \leq & \frac{1}{a_{\min}^2 K} \\ & \times \sum_{s_t \neq \tilde{s}_t} \Bigg(\operatorname{tr}\left\{W_{s_t \tilde{s}_t}^2 Q_{\tilde{s}_t}\right\} \cdot \left|R_\lambda\left(s_t, \tilde{s}_t\right)\right| \cdot \left|Q_{s_t}\right|^{-\lambda} \\ & \qquad \times \left|Q_{\tilde{s}_t}\right|^{\lambda-1} + \operatorname{tr}\left\{W_{s_t \tilde{s}_t}^2 R_\lambda\left(s_t, \tilde{s}_t\right)\right\} \Bigg), \\ & 0 < \lambda < 1. \end{aligned} \qquad (52)$$

It is worthwhile noting that our mse analysis follows the analysis in [18] but, the latter deals with a memoryless regime-switching process and a Toeplitz covariance matrix, whereas in our case both assumptions do not hold.

V. Model Estimation

In this section, we address the problem of estimating the model parameters $\phi \triangleq \{\boldsymbol{\pi}^{(0)}, A, \zeta_1, \ldots, \zeta_m, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m\}$. The ML estimation approach is commonly used for estimating the parameters of GARCH models (e.g., [1], [3], [28]) and also for estimating the transition probability matrices (e.g., [21]). The model parameters are estimated from a training data set of N clean signals of lengths $T_n, n = 1, \ldots, N$. Let $\left\{\mathbf{X}_t^{(n)}\right\}$ denote the spectral coefficients of the nth clean training signal and let $\mathcal{X}^{\tau,(n)} \triangleq \left\{\mathbf{X}_t^{(n)} | t = 0, \ldots, \tau\right\}$. The conditional distribution of the vector $\mathbf{X}_t^{(n)}$ given its past observations is a mixing of zero mean Gaussian vectors with diagonal covariance matrices $\hat{R}_{s_t}^{x,(n)}$

$$b\left(\mathbf{X}_t^{(n)} | \mathcal{X}^{t-1,(n)}\right) = \sum_{s_t} p\left(s_t | \mathcal{X}^{t-1,(n)}\right) \times b\left(\mathbf{X}_t^{(n)} | s_t, \hat{R}_{s_t}^{x,(n)}\right). \qquad (53)$$

Given a set of model parameters ϕ, the diagonal covariance matrix of the density $b\left(\mathbf{X}_t^{(n)} | s_t, R_{s_t}^{x,(n)}\right)$ can be recursively estimated by using the estimation algorithm introduced in Section III, where the signal observations are known in this case. Assuming that the process is asymptotically wide-sense stationary, and that the training sequences are sufficiently large, the initial state probabilities, $\boldsymbol{\pi}^{(0)}$, and the initial conditional variance, $\boldsymbol{\lambda}_{0|-1,s_0}$, have negligible contribution to the total likelihood. Therefore it is convenient to choose in the following optimization problem the stationary values as the initial values, i.e., $\hat{\boldsymbol{\lambda}}_{0k|-1,s_0} = \hat{\Phi}\hat{\boldsymbol{\zeta}}$ as the initial conditional variances, and $\boldsymbol{\pi}^{(0)} = \hat{\boldsymbol{\pi}}$ as the initial state probabilities.

The conditional log-likelihood of the training set is given by

$$\mathcal{L}(\phi) = \sum_n \sum_{t=0}^{T_n - 1} \log b\left(\mathbf{X}_t^{(n)} | \mathcal{X}^{t-1,(n)}; \phi\right). \qquad (54)$$

Using the constraints in (6) and imposing $\hat{A}$ to be a transition probability matrix, the ML estimates of the model parameters ϕ can be obtained by solving the following nonlinear constrained optimization problem:

$$\begin{aligned} & \max_{\hat{\phi}} \mathcal{L}(\phi) \\ & \quad \text{s.t. } \hat{\zeta}_i > 0, \quad \hat{\alpha}_i \geq 0, \ \hat{\beta}_i \geq 0, \\ & \qquad \sum_{j=1}^{M} \hat{a}_{ij} = 1 \quad \forall\, i \in \{1, \ldots, m\}. \end{aligned} \qquad (55)$$

For a given parameters set $\hat{\phi}$, the sequence of state dependent conditional variances $\left\{\hat{\Lambda}_t\right\}$ can be evaluated recursively according to the method described in Section III and so is the set of conditional state probabilities $p\left(s_t | \mathcal{X}^{t-1}\right)$. The conditional log-likelihood (54) can then numerically maximized under the linear constrains of (55) as specified in [3], [8] or by using sequential quadratic programming [29], [30]. The computational complexity required for the model estimation is much higher than that required for a single-regime GARCH model since m^2 parameters are to be estimated for the transition probabilities matrix and in addition $3m$ GARCH parameters are to be evaluated. However, using the Markov-switching model, the optimization problem needs to be solved only once, prior to the restoration procedure. It is well known that the optimal set of parameters, ϕ, is not necessarily unique in a Markovian model [21] and in addition, the numerical optimization solution may only guarantee a local maxima of the likelihood function. However, the flexibility of the model enables better results than that achievable with a single-regime GARCH model [3], [4]. This is also shown in our simulation results, both for MSTF-GARCH processes and for speech signals.

VI. Experimental Results

We demonstrate the performance of the proposed algorithm when applied to restoration of noisy MSTF-GARCH signals, and to estimation of conditional variances and squared absolute values of speech signals in the STFT domain.

A. MSTF-GARCH Signals

The proposed model estimation and signal restoration algorithm has been applied to MSTF-GARCH models of 3 and 5 regimes, degraded by additive independent white noise with 0 to 15 dB input signal-to-noise ratio (SNR). For each state space ($m = 3, 5$), a set of 20 stationary models have been simulated with uniformly distributed parameters on the interval $(0, 1]$. For each model, the parameters, ϕ, are estimated from a set of 10 training signals, each of time length $T = 100$ and dimension $K = 100$. The estimated parameters are employed for restoration of a set of test signals containing 20 noisy signals of the same size, and basically four types of estimated variances are compared by incorporating them into the signal's recursive mmse estimator of (17) and (18). The "*theoretical limit*" is referred to as the estimator which exploits the true conditional variances, $\boldsymbol{\lambda}_{t|t-1,s_t}$, of the simulated process. This estimator is the optimal estimator in the mmse sense and its performance is compared with those of the recursive estimators. The "*MSTF-GARCH, true model*" is referred to as the recursive signal estimator, described in Section III, which manipulates the true parameters set, ϕ, and the "*MSTF-GARCH,* $m = i$" estimator employs a set of estimated parameters, $\hat{\phi}$, assuming that the model has i regimes. For the "*MSTF-GARCH,* $m = i$" estimator, the set of parameters, $\hat{\phi}$, is estimated using the ML approach as described in Section V. The performance of our algorithm is also compared with that of an estimator that assumes a "*constant variance*" process. For that estimator (only), the vector of "stationary" variances, are evaluated for each noisy signal from the corresponding clean signal.

Fig. 1(a) shows the SNR improvement obtained by using the different estimators, when applied to 3-state MSTF-GARCH signals. It can be seen that even when assuming a small number of regimes, still the MSTF-GARCH estimator outperforms the "*constant variance*" estimator, and the results achieved by assuming 3 or 5 regimes are comparable to those obtained by using the true model parameters. Fig. 1(b) shows estimation results for 5-state MSTF-GARCH processes, under the assumption of 1, 3, 5, or 7 regimes. The estimation performances improve with the increase of the number of assumed regimes, but using a larger number of regimes than the true number (e.g., 7 instead of 5 or 5 instead of 3) yields less accurate results.

The time-varying behavior of the recursive estimator is demonstrated for a 5-state MSTF-GARCH signal degraded by additive white noise with 5 dB SNR. Fig. 2 shows trace of the instantaneous output SNR for each time frame, obtained by the optimal estimator, the recursive estimators with presumable 1 or 5 regimes (i.e., "*MSTF-GARCH,* $m = 1, 5$") and a "*constant variance*" estimator. The varying volatility of the process implies time-varying performances for all those estimators. Nevertheless, under the assumption of five regimes our recursive estimator follows the optimal estimator with a relatively small degradation in performance. The single-regime estimator yields comparable results as the 5-regimes estimator for frames with large input SNR. However, for frames with low input SNR the results obtained by the single-regime estimator are comparable to those obtained by the "*constant variance*" estimator.

B. Speech Signals

The idea of using different states for the enhancement of speech signals was first introduced by Drucker [31]. He assumed five categories of speech signals, comprising fricatives, stops, vowels, glides, and nasals. The application of HMMs for speech enhancement requires a higher number of states [16], [17] since these models allow only a single density, or a finite set of mixture-densities, for the spectral coefficients in each state. The GARCH-based models allow continuous values of conditional variances with possible transients resulting from switching states. Hence, a small number of states may be sufficient for the representation of the coefficients' second-order moments. Furthermore, the dynamic of the spectral coefficients

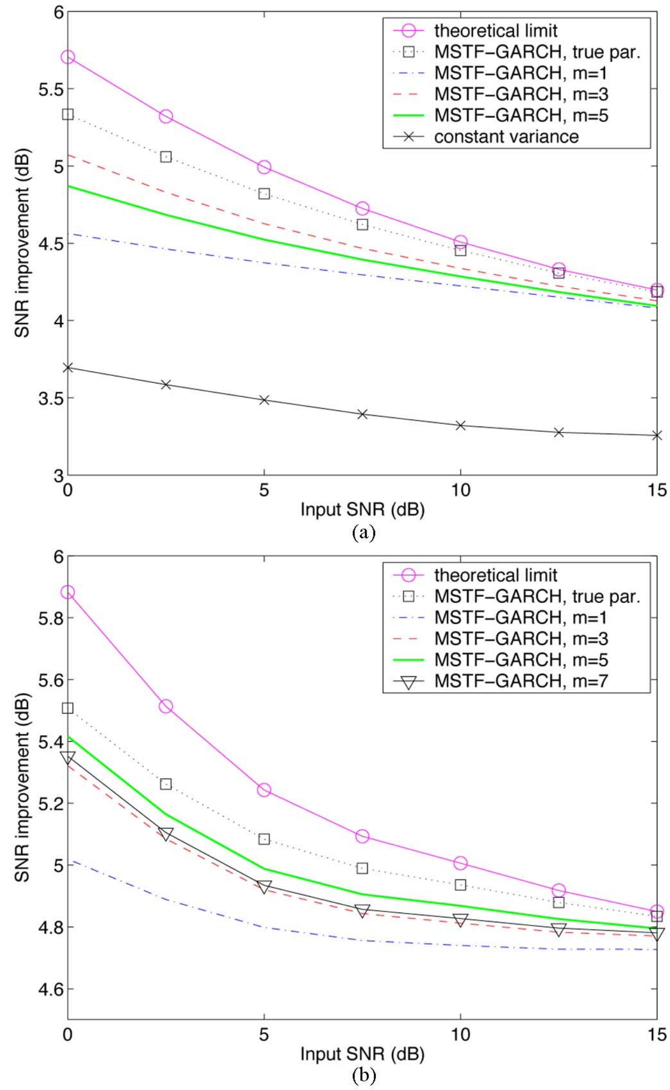


Fig. 1. SNR improvements obtained by using different MSTF-GARCH based estimators when applied to: (a) 3-state MSTF-GARCH signals; and (b) 5-state MSTF-GARCH signals. MSTF-GARCH models with various number of regimes are considered and compared with the true MSTF-GARCH parameters, the theoretical limit, and a constant variance estimator.

is frequency dependent. Therefore, we assume different parameters in different sub-bands.

The speech signals used in our evaluation are taken from the TIMIT database. The training set includes 10 different utterances from 10 different speakers, half male and half female. The speech signals are sampled at 8 kHz and normalized to the same energy. Transformation into the STFT domain is obtained by using half overlapping Hamming analysis window of 32 ms length. We consider 1, 3, and 5-state MSTF-GARCH models for the speech signals and estimate the one-frame-ahead conditional variance for test speech signals, not on the training set. Fig. 3 shows typical estimates of the one-frame-ahead conditional variance, $\hat{\lambda}_{tk|t-1}$, at frequencies of 1, 2, and 3 kHz, using the different MSTF-GARCH models and assuming independent model parameters in each frequency subband. The estimated conditional variances are compared with the clean signal's squared absolute value $|X_{tk}|^2$. It can be seen that by increasing the number of regimes, the conditional variance

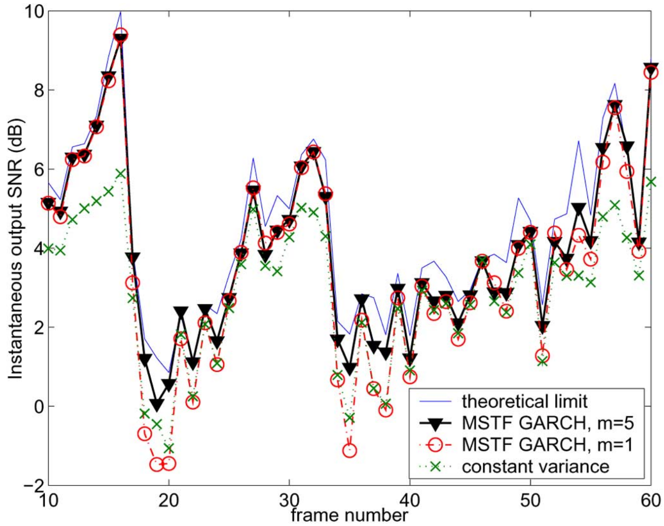


Fig. 2. Trace of instantaneous output SNR achieved by the proposed algorithm when applied to a realization of a 5-state MSTF-GARCH process degraded by additive white noise with 5 dB SNR, and restored by an MSTF-GARCH estimator, assuming 1 and 5 states.

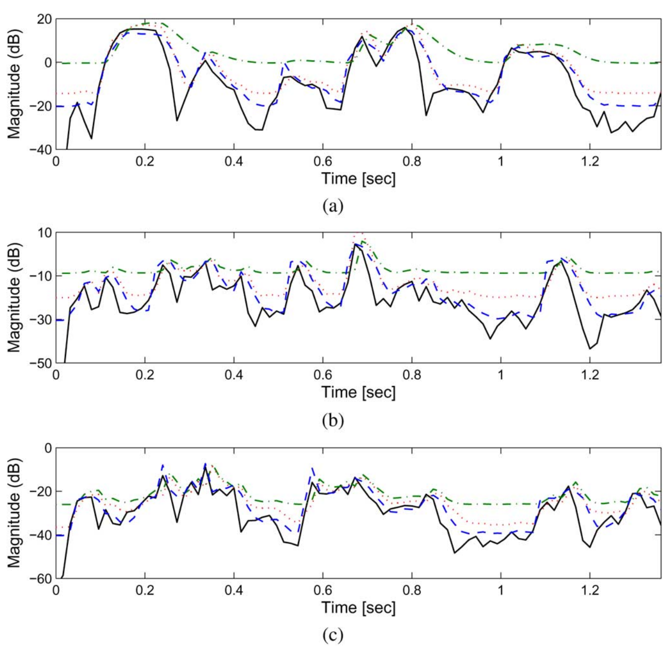


Fig. 3. Typical traces of one-frame-ahead conditional variance estimates for speech signals at frequencies: (a) 1 kHz; (b) 2 kHz; and (c) 3 kHz. The conditional variances are estimated by MSTF-GARCH models of single-state (dashed-dotted line), 3 states (dotted line), and 5 states (dashed line), and compared with the clean signal's squared absolute value (solid line).

yields a better prediction of the squared absolute value of the signal. Moreover, it can be seen that the conditional variance estimated by a single-regime model is smoother than that estimated based on a multiregime model, and the latter better tracks rapid changes in the signal's energy with possible switching of regimes. During the first few frames, the speech signal is absent and thus, as long as the squared absolute value is below the minimum variance allowed by the model, the predicted variances are determined by the model threshold. However, the predicted variances converge to the absolute squared value as soon as the latter exceeds this threshold. Larger number of states may allow better representation of the conditional variance in different magnitude ranges and different volatilities, at the expense of greater computational com

Many speech enhancement algorithms employ the decision-directed approach for the speech spectral variance estimation [27], [32]. Accordingly,

$$\hat{\lambda}_{tk}^{DD} = \max\left\{\bar{\alpha}\left|\hat{X}_{t-1,k}\right|^2 + (1-\bar{\alpha}) \times \left(|Y_{tk}|^2 - \sigma_k^2\right), \xi_{\min}\sigma_k^2\right\} \quad (56)$$

where $\bar{\alpha}(0 \leq \bar{\alpha} \leq 1)$ is a weighting factor that controls the tradeoff between noise reduction and transient distortion introduced into the signal. A larger value of $\bar{\alpha}$ results in a greater reduction of the musical noise phenomena, but at the expense of attenuated speech onsets and audible modifications of transient components. The parameter $\xi_{\min}$ is a lower bound on the *a priori* SNR.

The GARCH modeling enables an analytical derivation of the decision-directed estimator [33]. Considering the degenerated case of a single-state and a single-frequency ARCH (1) model (i.e., $\beta = 0$), the update step (22) can be written as

$$\hat{\lambda}_{tk|t} = \bar{\alpha}_{tk}\hat{\lambda}_{tk|t-1} + (1-\bar{\alpha}_{tk})\left(|Y_{tk}|^2 - \sigma_k^2\right) \quad (57)$$

with

$$\bar{\alpha}_{tk} \triangleq 1 - \frac{\hat{\lambda}_{tk|t-1}^2}{\left(\hat{\lambda}_{tk|t-1} + \sigma_k^2\right)^2} \quad (58)$$

and $0 < \bar{\alpha}_{tk} < 1$. Substituting the propagation step for $\hat{\lambda}_{tk|t-1}$ (23) into (57) with $\alpha = 1$, we obtain

$$\hat{\lambda}_{tk|t} = \bar{\alpha}_{tk} E\left\{|X_{t-1,k}|^2 \,|\mathcal{Y}^{t-1}\right\} + (1-\bar{\alpha}_{tk})\left(|Y_{tk}|^2 - \sigma_k^2\right) + \bar{\alpha}_{tk}\zeta. \quad (59)$$

For $\zeta \ll E\left\{|X_{t-1,k}|^2 \,|\mathcal{Y}^{t-1}\right\}$, (59) is similar to the decision-directed variance estimation (56) with $\bar{\alpha}_{tk} \equiv \bar{\alpha}$ and where $E\left\{|X_{t-1,k}|^2 \,|\mathcal{Y}^{t-1}\right\}$ holds for $\left|\hat{X}_{t-1,k}\right|^2$ which is the squared absolute value of the spectral coefficient estimate based on the observations $\mathcal{Y}^{t-1}$. Accordingly, the degenerated ARCH-based variance estimation with $\alpha = 1$ and low valued ζ is closely related to the decision-directed estimator with a time-varying frequency-dependent weighting factor $\bar{\alpha}_{tk}$. However, the GARCH (and ARCH) modeling approach manipulates the spectral variance as a random process, whereas the decision-directed approach assumes the spectral variance is a parameter which is heuristically evaluated. In addition, the decision-directed approach thresholds the estimated variance to be larger than $\xi_{\min}\sigma_k^2$ while in the GARCH modeling, the lower bound is inherently incorporated into the variance estimation.

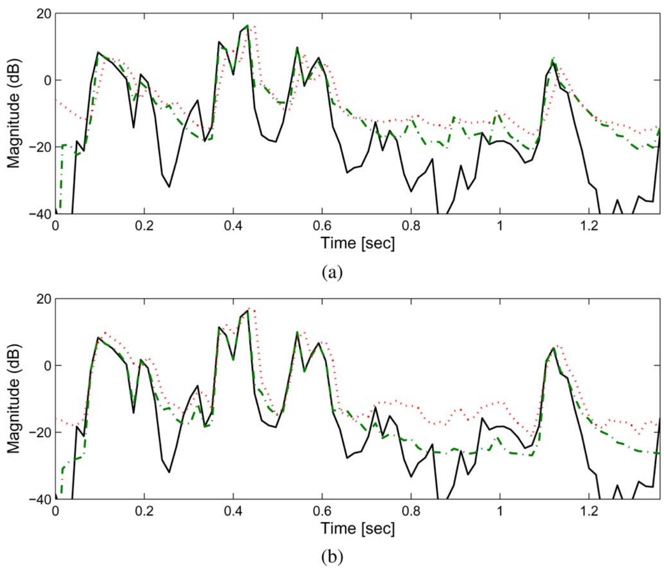


Fig. 4. Typical traces of estimated squared absolute values for speech signal at frequency of 2 kHz. The variances are estimated by a 5-state MSTF-GARCH model (dashed-dotted line), decision-directed approach (dotted line) and compared with the clean signal's squared absolute value (solid line). The SNRs are (a) 0 dB and (b) 10 dB.

Since $\hat{\lambda}_{tk|t-1} > \zeta$, from (22) we obtain the following lower bound

$$\hat{\lambda}_{tk|t} > \frac{\hat{\lambda}_{tk|t-1}}{\hat{\lambda}_{tk|t-1} + \sigma_k^2} \left(\sigma_k^2 + \frac{\hat{\lambda}_{tk|t-1}}{\hat{\lambda}_{tk|t-1} + \sigma_k^2} |Y_{tk}|^2 \right) \zeta > 0. \tag{60}$$

Modeling the spectral coefficients as an MSTF-GARCH allows further flexibility for the variance estimation. Fig. 4 demonstrates the estimated squared absolute values of a speech signal corrupted by a white Gaussian noise with SNR of (a) 0 dB and (b) 10 dB. The signal squared absolute value at frequency of 2 kHz is compared with its estimated variance using 5-state MSTF-GARCH model and by using the decision-directed approach. It shows that the MSTF-GARCH approach with five states yields a better estimate of the squared absolute value both under high and low SNR conditions, especially in low energy bins. Furthermore, the MSTF-GARCH approach enables a better tracking of rapid changes in the coefficients energy than the decision-directed approach.

The differences between Figs. 3 and 4 is that the former demonstrates the *prediction* of the coefficients' variances (i.e., the conditional variance) in a noiseless environment while the latter shows their second-order moments' estimation in a noisy environment. The variance prediction has a small delay of tracking rapid changes and the update step yields a better estimate of the squared absolute value in high energy bins. However, when noisy observations are employed, low-energy bins may be under the noise level and thus the estimation may be less accurate (for both the MSTF-GARCH approach and the decision-directed approach).

Figs. 3 and 4 demonstrate that the proposed MSTF-GARCH model, when compared to a single-regime model, or to the decision-directed approach, improves the variance prediction and the squared absolute value estimation of speech signals in the STFT domain. Still, one needs to derive a frequency-dependent model and to estimate the signal presence probability in each time-frequency bin of the noisy speech signal based on the proposed model, which is a subject for further research.

VII. Conclusion

We have proposed a statistical model for nonstationary processes with time-varying volatility structure in the STFT domain. Exploiting the advantages of both the conditional heteroscedasticity structure of GARCH models and the time-varying characteristics of hidden Markov chains, we model the expansion coefficients as multivariate, complex GARCH process with Markov-switching regimes. The correlation between successive coefficients in the time-frequency domain is taken into consideration by using the GARCH formulation which specifies the conditional variance as a linear function of its past values and past squared innovations. The time-varying structure of the conditional variance is determined by a hidden Markov chain which allows a different GARCH formulation in each state.

We showed that an ML estimate of the model can be practically obtained from training signals (assuming that the number of states is known), and developed a recursive algorithm for estimating the signal and its conditional variance in the STFT domain from its noisy observations. The conditional variance is recursively estimated for any regime by iterating propagation and update steps, while the evaluation of the regime conditional probabilities is based on the recursive correlation of the process. Experimental results demonstrate the improved performance of the proposed recursive algorithm compared to using an estimator which assumes a stationary process, even when the number of assumed regimes is smaller than the true number. When the number of assumed regimes approaches the true one, the recursive estimator yields comparable restoration results to those achievable by using the true model parameters. The conditional variance of an MSTF-GARCH process, as well as the instantaneous SNR on each frame, change over time. It is demonstrated that the recursive estimation approach has relatively small performance degradation compared to the theoretical estimation limit in the mmse sense. Performance evaluation with real speech signals demonstrates better variance estimation when using a multiregime model, compared to using a single-regime model, and improved squared absolute value estimation in a noisy environment compared to using the decision-directed approach.

Several extensions of this paper, which may be interesting for further research, include analysis of the algorithm sensitivity to the number of the assumed states, the parameters values and the training set; generalization of the multivariate complex Markov-switching GARCH model, such that the conditional covariance matrix is not necessarily diagonal and the correlation between distinct frequency-bins is also taken into account; and finally estimation of the signal presence probability in the time-frequency domain and modification of the recursive signal estimation algorithm under signal presence uncertainty.

Acknowledgment

The authors thank the anonymous reviewers for their valuable comments and helpful suggestions.

REFERENCES

[1] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Economet.*, vol. 31, no. 3, pp. 307–327, 1986.

[2] R. F. Engle, "Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, Jul. 1982.

[3] F. Klaassen, "Improving GARCH volatility forecasts with regime-switching GARCH," *Empir. Econom.*, vol. 27, no. 2, pp. 363–394, Mar. 2002.

[4] M. Haas, S. Mittnik, and M. S. Paolella, "A new approach to Markov-switching GARCH models," *J. Financial Economet.*, vol. 2, no. 4, pp. 493–530, 2004.

[5] J. Marcucci, "Forecasting stock market volatility with regime-switching GARCH models," *Studies Nonlin. Dynam. Economet.*, vol. 9, no. 4, 2005, Article 6.

[6] J. Cai, "A Markov model of switching-regime ARCH," *J. Business Econom. Statist.*, vol. 12, no. 3, pp. 309–316, Jul. 1994.

[7] J. D. Hamilton and R. Susmel, "Autoregressive conditional heteroskedasticity and changes in regime," *J. Economet.*, vol. 64, pp. 307–333, Jul. 1994.

[8] S. F. Gray, "Modeling the conditional distribution of interest rates as a regime-switching process," *J. Finan. Econom.*, vol. 42, pp. 27–62, Sep. 1996.

[9] I. Cohen, "Modeling speech signals in time-frequency domain using GARCH," *Signal Process.*, vol. 84, no. 12, pp. 2453–2459, Dec. 2004.

[10] I. Cohen, "Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models," *Signal Process.*, vol. 68, no. 4, pp. 698–709, Apr. 2006.

[11] I. Cohen, "From volatility modeling of financial time-series to stochastic modeling and enhancement of speech signals," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. New York: Springer, 2005, ch. 5, pp. 97–114.

[12] M. Abdolahi and H. Amindavar, "GARCH coefficients as feature for speech recognition in persian isolated digit," in *Proc. 30th IEEE Inte. Conf. Acoust. Speech Signal Process. (ICASSP)*, Philadelphia, PA, May 2005, vol. 17–21, pp. I.957–I.960.

[13] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition," *Bell Syst. Tech. J.*, vol. 62, no. 4, pp. 1035–1074, Apr. 1983.

[14] N. Z. Tishby, "On the application of mixture AR hidden Markov models to text independent speaker recognition," *IEEE Trans. Signal Process.*, vol. 39, no. 3, pp. 563–570, Mar. 1991.

[15] Y. Ephraim and W. J. Roberts, "Revisiting autoregressive hidden Markov modeling of speech signals," *IEEE Signal Process. Lett.*, vol. 12, pp. 166–169, Feb. 2005.

[16] Y. Ephraim, D. Malah, and B.-H. Juang, "On the application of hidden Markov models for enhancing noisy speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1846–1856, Dec. 1989.

[17] Y. Ephraim, "A bayesian estimation approach for speech enhancement using hidden marks models," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 725–735, Apr. 1992.

[18] Y. Ephraim and N. Merhav, "Lower and upper bounds on the minimum mean-square error in composite source signal estimation," *IEEE Trans. Inf. Theory*, vol. 38, no. 6, pp. 1709–1724, Nov. 1992.

[19] A. Abramson and I. Cohen, "On the stationarity of GARCH processes with Markov switching regimes," *Econometric Theory*, vol. 23, no. 3, pp. 485–500, 2007.

[20] A. Abramson and I. Cohen, "Asymptotic stationarity of Markov-switching time-frequency GARCH processes," in *Proc. 30th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2006, pp. III 452–445.

[21] Y. Ephraim and N. Merhav, "Hidden marks processes," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1518–1569, Jun. 2002.

[22] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, pp. 504–512, 2001.

[23] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary environments," *Signal Process.*, vol. 81, pp. 2403–2418, Nov. 2001.

[24] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, pp. 466–475, Sep. 2003.

[25] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[26] A. Abramson and I. Cohen, "State smoothing in Markov-switching time-frequency GARCH models," *IEEE Signal Process. Lett.*, vol. 13, pp. 377–380, Jun. 2006.

[27] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[28] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.

[29] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. New York: Academic, 1981.

[30] S. Han, "A globally convergent method for nonlinear programming," *J. Optimiz. Theory Applicat.*, vol. 22, no. 4, pp. 297–309, 1977.

[31] H. Drucker, "Speech processing in a high ambient noise environment," *IEEE Trans. Audio Electroacoust.*, vol. AU-16, no. 2, pp. 165–168, Jun. 1968.

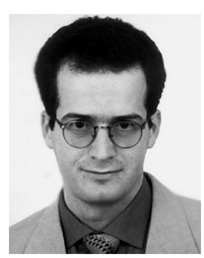
[32] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech Audio Process.*, vol. 2, pp. 345–349, Apr. 1994.

[33] I. Cohen, "Relaxed statistical model for speech enhancement and *a priori* SNR estimation," *IEEE Trans. Speech Audio Process.*, vol. 13, pp. 870–881, Sep. 2005.

Ari Abramson (S'06) received the B.Sc. degree in electrical engineering from Tel-Aviv University, Israel, in 2002.

He is currently pursuing the Ph.D. degree in electrical engineering in the direct-tract doctoral program of the Technion—Israel Institute of Technology, Haifa, Israel. From 1993 to 2004, he served as a combat copilot in the Israeli Air-Force, and since 2004 he has been a flight-test engineer in reserve duty. His research interests are statistical signal processing, speech enhancement, and detection and estimation theory.

Mr. Abramson received the Wolf foundation excellence award in 2005 and the Best Student Paper Award at the International Workshop on Acoustic, Echo and Noise Control in 2006.



Israel Cohen (M'01–SM'03) received the B.Sc. (*summa cum laude*), M.Sc., and Ph.D. degrees in electrical engineering in 1990, 1993, and 1998, respectively, all from the Technion—Israel Institute of Technology, Haifa, Israel.

From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT. In 2001, he joined the Electrical Engineering Department of the Technion, where he is currently an Associate Professor. His research interests are statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification, and adaptive filtering.

Dr. Cohen received in 2005 and 2006 the Technion Excellent Lecturer awards. He serves as Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and the IEEE SIGNAL PROCESSING LETTERS. He was a Guest Editor of a Special Issue of the *EURASIP Journal on Applied Signal Processing* on Advances in Multimicrophone Speech Processing, as well as a Special Issue of the *EURASIP Speech Communication Journal* on Speech Enhancement. He is a Coeditor of the Multichannel Speech Processing section of the *Springer Handbook of Speech Processing*.