Modeling and Identification of Nonlinear Systems in the Short-Time Fourier Transform Domain

Yekutiel Avargel, *Student Member, IEEE* and Israel Cohen, *Senior Member, IEEE*

Abstract

In this paper, we introduce a novel approach for improved nonlinear system identification in the short-time Fourier transform (STFT) domain. We first derive explicit representations of discrete-time Volterra filters in the STFT domain. Based on these representations, approximate nonlinear STFT models, which consist of parallel combinations of linear and nonlinear components, are developed. The linear components are represented by crossband filters between subbands, while the nonlinear components are modeled by multiplicative cross-terms. We consider the identification of quadratically nonlinear systems and show that a significant reduction in computational cost as well as substantial improvement in estimation accuracy can be achieved over a time-domain Volterra model, particularly when long-memory systems are considered. Experimental results validate the theoretical derivations and demonstrate the effectiveness of the proposed approach.

Index Terms

Nonlinear systems, Volterra filters, system identification, subband filtering, short-time Fourier transform, time-frequency analysis.

I. Introduction

Identification of linear systems has been studied extensively and is of major importance in diverse fields of signal processing [1], [2]. However, in many real-world applications, the considered systems

This research was supported by the Israel Science Foundation (grant no. 1085/05).

The authors are with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel. E-mail addresses: kutiav@tx.technion.ac.il (Y. Avargel), icohen@ee.technion.ac.il (I. Cohen); tel.: +972-4-8294731; fax: +972-4-8295757.

exhibit certain nonlinearities that cannot be sufficiently estimated by conventional linear models. Examples of such applications include acoustic echo cancellation [3]–[5], channel equalization [6], [7], biological system modeling [8], image processing [9], and loudspeaker linearization [10]. Volterra filters [11]–[16] are widely used for modeling nonlinear physical systems, such as loudspeaker-enclosure-microphone (LEM) systems in nonlinear acoustic echo cancellation applications [4], [17], [18], and digital communication channels [6], [19], just to mention a few. An important property of Volterra filters, which makes them useful in nonlinear estimation problems, is the linear relation between the system output and the filter coefficients. Many approaches, which attempt to estimate the Volterra kernels in the time domain, employ conventional linear estimation methods in batch (e.g., [15], [20]) or adaptive forms (e.g., [4], [21]). A common difficulty associated with time-domain methods is their high computational cost, which is attributable to the large number of parameters of the Volterra model. This problem becomes even more crucial when estimating systems with relatively large memory length, as in acoustic echo cancellation applications. Another major drawback of the Volterra model is its severe ill-conditioning [22], which leads to high estimation-error variance and to slow convergence of the adaptive Volterra filter. To overcome these problems, several approximations for the time-domain Volterra filter have been proposed, including orthogonalized power filters [23], Hammerstein models [24], parallel-cascade structures [25], and multi-memory decomposition [26].

Alternatively, frequency-domain methods have been introduced for Volterra system identification, aiming at estimating the so-called Volterra transfer functions [27]–[29]. Statistical approaches based on higher order statistics (HOS) of the input signal use cumulants and polyspectra information [27]. These approaches have relatively low computational cost, but often assume a Gaussian input signal, which limits their applicability. In [28] and [29], a discrete frequency-domain model is defined, which approximates the Volterra filter in the frequency domain using multiplicative terms. Although this approach assumes no particular statistics for the input signal, it requires a long duration of the input signal to validate the multiplicative approximation and to achieve satisfactory performance. When the data is of limited size (or when the nonlinear system is not time-invariant), this long duration assumption is very restrictive.

In this paper, we introduce a novel approach for improved nonlinear system identification in the short-time Fourier transform (STFT) domain, which is based on a time-frequency representation of the Volterra filter. A typical nonlinear system identification scheme in the STFT domain is illustrated in Fig. 1. Similarly to STFT-based linear identification techniques [30]–[32], representing and identifying nonlinear systems in the STFT domain is motivated by a reduction in computational cost compared to time-domain methods, due to processing in distinct subbands. Together with a reduction in the spectral dynamic range

of the input signal, the reduced complexity may also lead to a faster convergence of nonlinear adaptive algorithms. Consequently, a proper model in the STFT domain may facilitate a practical alternative for conventional nonlinear models, especially in estimating nonlinear systems with relatively long memory, which cannot be practically estimated by existing methods. We show that a homogeneous time-domain Volterra filter [11] with a certain kernel can be perfectly represented in the STFT domain, at each frequency bin, by a sum of Volterra-like expansions with smaller-sized kernels. This representation, however, is impractical for identifying nonlinear systems due to the extremely large complexity of the model. We develop an approximate nonlinear model, which simplifies the STFT representation of Volterra filters and significantly reduces the model complexity. The resulting model consists of a parallel combination of linear and nonlinear components. The linear component is represented by crossband filters between the subbands [30], [33], while the nonlinear component is modeled by multiplicative cross-terms, extending the so-called cross-multiplicative transfer function (CMTF) approximation [34]. It is shown that the proposed STFT model generalizes the conventional discrete frequency-domain model [28], and forms a much richer representation for nonlinear systems. Concerning system identification, we employ the proposed model and introduce an off-line scheme for estimating the model parameters using a least-squares (LS) criterion. The proposed approach is more advantageous in terms of computational complexity than the time-domain Volterra approach. When estimating long-memory systems, a substantial improvement in estimation accuracy over the Volterra model can be achieved, especially for high signal-to-noise ratio (SNR) conditions. Experimental results with white Gaussian signals and real speech signals demonstrate the advantages of the proposed approach.

The paper is organized as follows. In Section II, we derive an explicit representation of discrete-time Volterra filters in the STFT domain. In Section III, we introduce a simplified model for nonlinear systems in the STFT domain. In Section IV, we consider off-line estimation of the proposed-model parameters and compare its complexity to that of the conventional time-domain approach. Finally, in Section V, we present some experimental results.

II. Representation of Volterra Filters in the STFT Domain

In this section, we represent discrete-time Volterra filters in the STFT domain. We first consider the quadratic case, and subsequently generalize the results to higher orders of nonlinearity. We show that a time-domain Volterra kernel can be perfectly represented in the STFT domain by a sum of smaller-sized kernels in each frequency bin. Throughout this work, unless explicitly noted, the summation indices range from $-\infty$ to ∞.

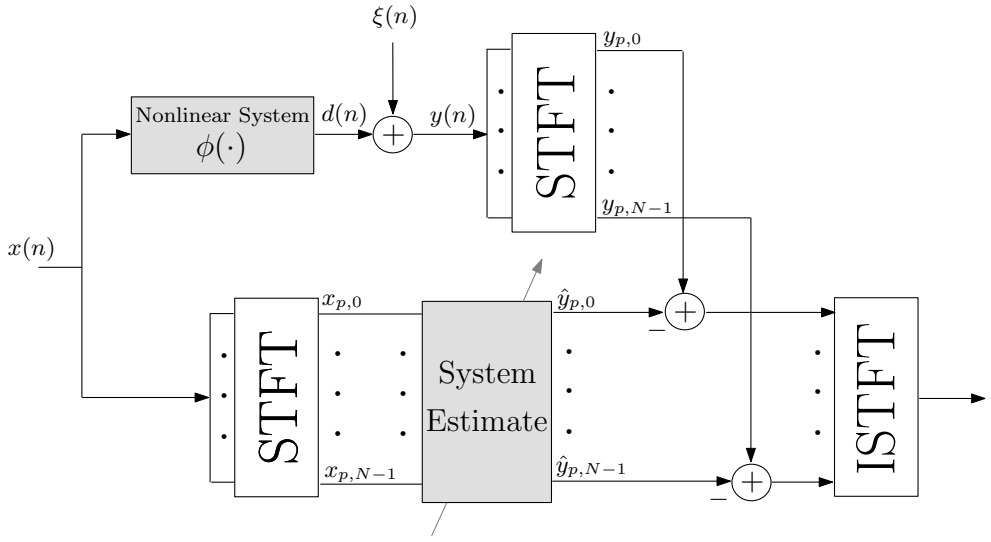

Fig. 1: Nonlinear system identification in the STFT domain. The unknown time-domain nonlinear system $\phi(\cdot)$ is estimated using a given model in the STFT domain.

A. Quadratically Nonlinear Systems

Consider a quadratically nonlinear system with an input $x(n)$ and an output $d(n)$. One of the most popular representations of such system is a second-order Volterra filter that relates $x(n)$ and $d(n)$ as follows:

$$\begin{aligned} d(n) &= \sum_{m=0}^{N_1-1} h_1(m)x(n-m) \\ &\quad + \sum_{m=0}^{N_2-1}\sum_{\ell=0}^{N_2-1} h_2(m,\ell)x(n-m)x(n-\ell) \\ &\triangleq d_1(n) + d_2(n)\,, \end{aligned} \tag{1}$$

where $h_1(m)$ and $h_2(m,\ell)$ are the linear and quadratic Volterra kernels, respectively, and $d_1(n)$ and $d_2(n)$ denote the corresponding output signals of the linear and quadratic homogeneous components. To find a representation of $d(n)$ in the STFT domain, let us first briefly review some definitions of the STFT representation of digital signals (for further details, see e.g., [35]).

The STFT representation of a signal $x(n)$ is given by

$$x_{p,k} = \sum_m x(m)\tilde{\psi}^*_{p,k}(m)\,, \tag{2}$$

where

$$\tilde{\psi}_{p,k}(n) \triangleq \tilde{\psi}(n-pL)e^{j\frac{2\pi}{N}k(n-pL)} \tag{3}$$

denotes a translated and modulated window function, $\tilde{\psi}(n)$ is an analysis window of length N, p is the frame index, k represents the frequency-bin index ($0 \leq k \leq N-1$), L is the translation factor (or the decimation factor, in filter-bank interpretation) and * denotes complex conjugation. The inverse STFT, i.e., reconstruction of $x(n)$ from its STFT representation $x_{p,k}$, is given by

$$x(n) = \sum_{p} \sum_{k=0}^{N-1} x_{p,k} \psi_{p,k}(n)\,, \tag{4}$$

where

$$\psi_{p,k}(n) \triangleq \psi(n - pL) e^{j\frac{2\pi}{N}k(n-pL)}\,, \tag{5}$$

and $\psi(n)$ denotes a synthesis window of length N. Substituting (2) into (4), we obtain the so-called completeness condition:

$$\sum_{p} \psi(n - pL)\tilde{\psi}^*(n - pL) = \frac{1}{N} \qquad \text{for all } n\,. \tag{6}$$

Given analysis and synthesis windows that satisfy (6), a signal $x(n) \in \ell_2(\mathbb{Z})$ is guaranteed to be perfectly reconstructed from its STFT coefficients $x_{p,k}$. However, for $L \leq N$ and for a given synthesis window $\psi(n)$, there might be an infinite number of solutions to (6); therefore, the choice of the analysis window is generally not unique [36], [37].

Using the linearity of the STFT, $d(n)$ in (1) can be written in the time-frequency domain as

$$d_{p,k} = d_{1;p,k} + d_{2;p,k}\,, \tag{7}$$

where $d_{1;p,k}$ and $d_{2;p,k}$ are the STFT representations of $d_1(n)$ and $d_2(n)$, respectively. It is well known that in order to perfectly represent a linear system in the STFT domain, crossband filters between subbands are generally required [30], [33]. Therefore, the output of the linear component can be expressed in the STFT domain as

$$d_{1;p,k} = \sum_{k'=0}^{N-1} \sum_{p'=0}^{\bar{N}_1 - 1} x_{p-p',k'} h_{p',k,k'}\,, \tag{8}$$

where $h_{p,k,k'}$ denotes a crossband filter of length $\bar{N}_1 = \lceil (N_1 + N - 1)/L \rceil + \lceil N/L \rceil - 1$ from frequency bin k' to frequency bin k. These filters are used for canceling the aliasing effects caused by the subsampling factor L. The crossband filter $h_{p,k,k'}$ is related to the linear kernel $h_1(n)$ by [30]

$$h_{p,k,k'} = \{h_1(n) * \phi_{k,k'}(n)\}|_{n=pL} \tag{9}$$

where the discrete-time Fourier transform (DTFT) of $\phi_{k,k'}(n)$ with respect to the time index n is given by

$$\Phi_{k,k'}(\omega) = \sum_{n} \phi_{k,k'}(n) e^{-jn\omega} = \tilde{\Psi}^*\left(\omega - \frac{2\pi}{N}k\right) \Psi\left(\omega - \frac{2\pi}{N}k'\right)\,, \tag{10}$$

where $\tilde{\Psi}(\omega)$ and $\Psi(\omega)$ are the DTFT of $\tilde{\psi}(n)$ and $\psi(n)$, respectively. Note that the energy of the crossband filter from frequency bin k' to frequency bin k generally decreases as $|k-k'|$ increases, since the overlap between $\tilde{\Psi}\left(\omega-(2\pi/N)\,k\right)$ and $\Psi\left(\omega-(2\pi/N)\,k'\right)$ becomes smaller. Recently, we have investigated the influence of crossband filters on a linear system identifier implemented in the STFT domain [30]. We showed that increasing the number of crossband filters not necessarily implies a lower steady-state mse in subbands. In fact, the inclusion of more crossband filters in the identification process is preferable only when high SNR or long data are considered. As will be shown later, the same applies also when an additional nonlinear component is incorporated into the model.

The representation of the quadratic component's output $d_2(n)$ in the STFT domain can be derived in a similar manner to that of the linear component. Specifically, applying the STFT to $d_2(n)$ we may obtain after some manipulations (see Appendix A)

$$\begin{aligned} d_{2;p,k} &= \sum_{k',k''=0}^{N-1} \sum_{p',p''} x_{p',k'} x_{p'',k''} c_{p-p',p-p'',k,k',k''} \\ &= \sum_{k',k''=0}^{N-1} \sum_{p',p''} x_{p-p',k'} x_{p-p'',k''} c_{p',p'',k,k',k''} \,. \end{aligned} \tag{11}$$

where $c_{p-p',p-p'',k,k',k''}$ may be interpreted as a response of the quadratic system to a pair of impulses $\{\delta_{p-p',k-k'}, \delta_{p-p'',k-k''}\}$ in the time-frequency domain. Equation (11) indicates that for a given frequency-bin index k, the temporal signal $d_{2;p,k}$ consists of all possible interactions between pairs of input frequencies. The contribution of each frequency pair $\{k', k''|\, k', k'' \in \{0,\ldots,N-1\}\}$ to the output signal at frequency bin k is given as a Volterra-like expansion with $c_{p',p'',k,k',k''}$ being its quadratic kernel. The kernel $c_{p',p'',k,k',k''}$ in the time-frequency domain is related to the quadratic kernel $h_2(n,m)$ in the time domain by (see Appendix A)

$$c_{p',p'',k,k',k''} = \left\{h_2(n,m) * \phi_{k,k',k''}(n,m)\right\}\big|_{n=p'L,\ m=p''L} \tag{12}$$

where $*$ denotes a 2D convolution and

$$\phi_{k,k',k''}(n,m) \triangleq \sum_{\ell} \tilde{\psi}^*(\ell) e^{-j\frac{2\pi}{N}k\ell} \psi(n+\ell) e^{j\frac{2\pi}{N}k'(n+\ell)} \psi(m+\ell) e^{j\frac{2\pi}{N}k''(m+\ell)} \,. \tag{13}$$

Equation (13) implies that for fixed k, k' and k'', the quadratic kernel $c_{p',p'',k,k',k''}$ is noncausal with $\lceil N/L \rceil - 1$ noncausal coefficients in each variable (p' and p''). Note that crossband filters are also noncausal with the same number of noncausal coefficients [30]. Hence, for system identification, an artificial delay of $(\lceil N/L \rceil - 1)\,L$ can be applied to the system output signal $d(n)$ in order to consider a

noncausal response. It can also be seen from (13) that the memory length of each kernel is given by

$$\bar{N}_2 = \left\lceil \frac{N_2 + N - 1}{L} \right\rceil + \left\lceil \frac{N}{L} \right\rceil - 1 \,, \tag{14}$$

which is approximately L times lower than the memory length of the time-domain kernel $h_2(m,\ell)$. The support of $c_{p',p'',k,k',k''}$ is therefore given by $\mathcal{D}\times\mathcal{D}$ where $\mathcal{D} = [1 - \lceil N/L \rceil, \ldots, \lceil (N_2 + N - 1)/L \rceil - 1]$.

To give further insight into the basic properties of the quadratic STFT kernels $c_{p',p'',k,k',k''}$, we apply the 2D DTFT to $\phi_{k,k',k''}(n,m)$ with respect to the time indices n and m, and obtain

$$\Phi_{k,k',k''}(\omega,\eta) = \tilde{\Psi}^* \left(\omega + \eta - \frac{2\pi}{N}k\right) \Psi\left(\omega - \frac{2\pi}{N}k'\right) \Psi\left(\omega - \frac{2\pi}{N}k''\right) . \tag{15}$$

By taking $\Psi(\omega)$ and $\tilde{\Psi}(\omega)$ to be ideal low-pass filters with bandwidths π/N (i.e., $\Psi(\omega) = 0$ and $\tilde{\Psi}(\omega) = 0$ for $\omega \notin [-\pi/2N, \pi/2N]$), a perfect STFT representation of the quadratic time-domain kernel $h_2(n,m)$ can be achieved by utilizing only kernels of the form $c_{p',p'',k,k',(k-k') \bmod N}$, since in this case the product of $\Psi(\omega - (2\pi/N)k')$, $\Psi(\omega - (2\pi/N)k')$ and $\tilde{\Psi}^*(\omega + \eta - (2\pi/N)k)$ is identically zero for $k'' \neq (k - k') \bmod N$. Practically, the analysis and synthesis windows are not ideal and their bandwidths are greater than π/N, so $\phi_{k,k',(k-k') \bmod N}(n,m)$, and consequently $c_{p',p'',k,k',(k-k') \bmod N}$, are not zero. Nonetheless, one can observe from (15) that the energy of $\phi_{k,k',k''}(n,m)$ decreases as $|k'' - (k - k') \bmod N|$ increases, since the overlap between the translated window functions becomes smaller. As a result, not all kernels in the STFT domain should be considered in order to capture most of the energy of the STFT representation of $h_2(n,m)$. This is illustrated in Fig. 2, which shows the energy of $\phi_{k,k',k''}(n,m)$, defined as $E_{k,k'}(k'') \triangleq \sum_{n,m} |\phi_{k,k',k''}(n,m)|^2$, for $k = 1$, $k' = 0$ and $k'' \in \{(k - k' + i) \bmod N\}_{i=-10}^{10}$, as obtained by using rectangular, triangular and Hann synthesis windows of length $N = 256$. A corresponding minimum-energy analysis window that satisfies the completeness condition [36] for $L = 128$ (50% overlap) is also employed. The results confirm that the energy of $\phi_{k,k',k''}(n,m)$, for fixed k and k', is concentrated around the index $k'' = (k - k') \bmod N$.

As expected from (15), the number of useful quadratic kernels in each frequency bin is mainly determined by the spectral characteristics of the analysis and synthesis windows. That is, windows with a narrow mainlobe (e.g., a rectangular window) yield the sharpest decay, but suffer from wider energy distribution over k'' due to relatively high sidelobes energy. Smoother windows (e.g., Hann window), on the other hand, enable better energy concentration. For instance, utilizing a Hann window reduces the energy of $\phi_{k,k',k''}(n,m)$ for $k'' = (k - k' \pm 8) \bmod N$ by approximately 30 dB, when compared to using a rectangular window. These results will be used in the next section for deriving a useful model for nonlinear systems in the STFT domain.

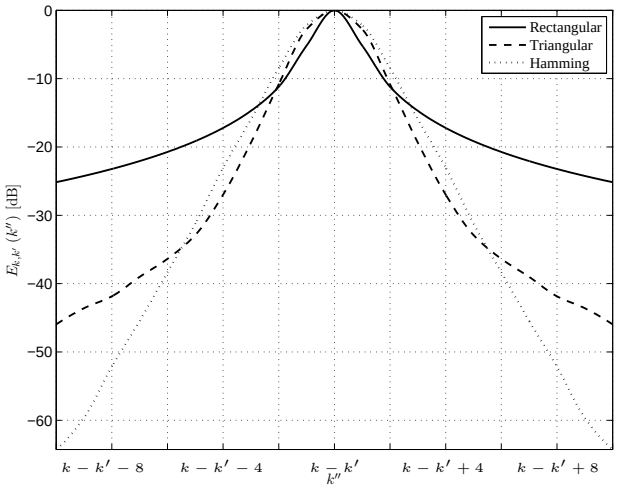

Fig. 2: Energy of $\phi_{k,k',k''}(n,m)$ [defined in (13)] for $k=1$ and $k'=0$, as obtained for different synthesis windows of length $N=256$.

B. High-Order Nonlinear Systems

Let us now consider a generalized qth-order nonlinear system with an input $x(n)$ and an output $d(n)$. A time-domain qth-order Volterra filter representation of this system is given by

$$d(n)=\sum_{\ell=1}^{q} d_\ell(n) \tag{16}$$

where $d_\ell(n)$ represents the output of the ℓth-order homogeneous Volterra filter, which is related to the input $x(n)$ by

$$d_\ell(n)=\sum_{m_1=0}^{N_\ell-1}\cdots\sum_{m_\ell=0}^{N_\ell-1} h_\ell(m_1,\dots m_\ell)\prod_{i=1}^{\ell}x(n-m_i) \tag{17}$$

where $h_\ell(m_1,\dots m_\ell)$ is the ℓth-order Volterra kernel, and N_ℓ ($1\le\ell\le q$) represents its memory length. This representation is called symmetric if the Volterra kernels satisfy [11]

$$h_\ell(m_1,\dots m_\ell)=h_\ell(m_{\pi(1)},\dots m_{\pi(\ell)}) \tag{18}$$

for any permutation $\pi(1,\dots,\ell)$. In order to reduce the redundancy of the symmetric representation, the triangular or regular representations may be employed (for further details, see e.g., [11]).

Applying the STFT to $d_\ell(n)$ and following a similar derivation to that made for the quadratic case [see (11)-(13), and Appendix A], we obtain after some manipulations

$$d_{\ell;p,k}=\sum_{k_1,\dots k_\ell=0}^{N-1}\sum_{p_1,\dots p_\ell} c_{p_1,\dots p_\ell,k,k_1,\dots k_\ell}\prod_{i=1}^{\ell}x_{p-p_i,k_i}\,. \tag{19}$$

Equation (19) implies that the output of an ℓth-order homogeneous Volterra filter in the STFT domain, at a given frequency-bin index k, consists of all possible combinations of ℓ input frequencies. The contribution of each ℓ-fold frequency indices $\{k_1, \dots k_\ell\}$ to the kth frequency bin is expressed in terms of an ℓth-order homogeneous Volterra expansion with the kernel $c_{p_1,\dots p_\ell,k,k_1,\dots k_\ell}$. Similarly to the quadratic case, it can be shown that the STFT kernel $c_{p_1,\dots p_\ell,k,k_1,\dots k_\ell}$ in the time-frequency domain is related to the kernel $h_\ell(m_1, \dots m_\ell)$ in the time domain by

$$c_{p_1,\dots p_\ell,k,k_1,\dots k_\ell} = \{h_\ell(m_1, \dots m_\ell) * \phi_{k,k_1,\dots k_\ell}(m_1, \dots m_\ell)\}|_{m_i=p_iL;\ i=1,\dots \ell.} \tag{20}$$

where $*$ denotes an ℓ-D convolution and

$$\phi_{k,k_1,\dots k_\ell}(m_1, \dots m_\ell) \triangleq \sum_n \tilde{\psi}^*(n) e^{-j\frac{2\pi}{N}kn} \prod_{i=1}^{\ell} \psi(m_i + n) e^{j\frac{2\pi}{N}k_i(m_i+n)}\,. \tag{21}$$

Equations (20)-(21) imply that for fixed indices $\{k_i\}_{i=1}^{\ell}$, the kernel $c_{p_1,\dots p_\ell,k,k_1,\dots k_\ell}$ is noncausal with $\lceil N/L \rceil - 1$ noncausal coefficients in each variable $\{p_i\}_{i=1}^{\ell}$, and its overall memory length is given by

$$\bar{N}_\ell = \left\lceil \frac{N_\ell + N - 1}{L} \right\rceil + \left\lceil \frac{N}{L} \right\rceil - 1\,. \tag{22}$$

Note that for $\ell = 1$ and $\ell = 2$, (19)-(21) reduce to the STFT representation of the linear kernel (8) and the quadratic kernel (11), respectively. Furthermore, applying the ℓ-D DTFT to $\phi_{k,k_1,\dots k_\ell}(m_1, \dots m_\ell)$ with respect to the time indices $m_1, \dots m_\ell$, we obtain

$$\Phi_{k,k_1,\dots k_\ell}(\omega_1, \dots \omega_\ell) = \tilde{\Psi}^* \left(\sum_{i=1}^{\ell} \omega_i - \frac{2\pi}{N} k \right) \prod_{m=1}^{\ell} \Psi \left(\omega_m - \frac{2\pi}{N} k_m \right)\,. \tag{23}$$

If both $\tilde{\Psi}(\omega)$ and $\Psi(\omega)$ were ideal low-pass filters with bandwidth $2\pi/\left(\lceil (\ell+1)/2 \rceil\, N\right)$, the overlap between the translated window functions in (23) would have been identically zero for $k_\ell \neq \left(k - \sum_{i=1}^{\ell-1} k_i\right) \bmod N$, and thus only kernels of the form $c_{p_1,\dots p_\ell,k,k_1,\dots k_\ell}$ where $k_\ell = \left(k - \sum_{i=1}^{\ell-1} k_i\right) \bmod N$ would have contributed to the output at frequency-bin index k. Practically, the energy is distributed over all kernels and particularly concentrated around the index $k_\ell = \left(k - \sum_{i=1}^{\ell-1} k_i\right) \bmod N$, as was demonstrated in Fig. 2 for the quadratic case ($\ell = 2$).

III. An Approximate Model for Nonlinear Systems in the STFT Domain

Representation of Volterra filters in the STFT domain involves a large number of parameters and high error variance, particularly when estimating the system from short and noisy data. In this section, we introduce an approximate model for improved nonlinear system identification in the STFT domain, which simplifies the STFT representation of Volterra filters and reduces the model complexity.

We start with an STFT representation of a second-order Volterra filter. Recall that modeling the linear kernel requires N crossband filters in each frequency bin [see (8)], where the length of each filter is approximately N_1/L. For system identification, however, only a few crossband filters need to be considered [30], which leads to a computationally efficient representation of the linear component. The quadratic Volterra kernel representation, on the other hand, consists of N^2 kernels in each frequency bin [see (11)], where the size of each kernel in the STFT domain is approximately $N_2/L \times N_2/L$. A perfect representation of the quadratic kernel is then achieved by employing $(NN_2/L)^2$ parameters in each frequency bin. Even though it may be reduced by considering the symmetric properties of the kernels, the complexity of such a model remains extremely large.

To reduce the complexity of the quadratic model in the STFT domain, let us assume that the analysis and synthesis filters are selective enough, such that according to Fig. 2, most of the energy of a quadratic kernel $c_{p',p'',k,k',k''}$ (for fixed k and k') is concentrated in a small region around the index $k'' = (k - k') \bmod N$. Accordingly, (11) can be efficiently approximated by

$$d_{2;p,k} \approx \sum_{\substack{k',k''=0 \\ (k'+k'') \bmod N = k}}^{N-1} \sum_{p',p''} x_{p-p',k'} x_{p-p'',k''} c_{p',p'',k,k',k''} \,. \tag{24}$$

A further simplification can be made by extending the so-called cross-multiplicative transfer function (CMTF) approximation, which was first introduced in [34], [38] for the representation of linear systems in the STFT domain. According to this model, a linear system is represented in the STFT domain by cross-multiplicative terms, rather than crossband filters, between distinct subbands. Following a similar reasoning, a kernel $c_{p',p'',k,k',k''}$ in (24) may be approximated as purely multiplicative in the STFT domain, so that (24) degenerates to

$$d_{2;p,k} \approx \sum_{\substack{k',k''=0 \\ (k'+k'') \bmod N = k}}^{N-1} x_{p,k'} x_{p,k''} c_{k',k''} \,. \tag{25}$$

We refer to $c_{k',k''}$ as a *quadratic cross-term*. The constraint $(k' + k'') \bmod N = k$ on the summation indices in (25) indicates that only frequency indices $\{k', k''\}$, whose sum is k or $k + N$[1], contribute to the output at frequency bin k. This concept is well illustrated in Fig. 3, which shows the (k', k'') two-dimensional plane. For calculating $d_{2;p,k}$ at frequency bin k, only points on the lines $k' + k'' = k$ and $k' + k'' = k + N$ need to be considered. Moreover, the quadratic cross-terms $c_{k',k''}$ have unique

[1]Since k and k' range from 0 to $N - 1$, the contribution of the difference interaction of two frequencies to the kth frequency bin corresponds to the sum interaction of the same two frequencies to the $(k + N)$th frequency bin.

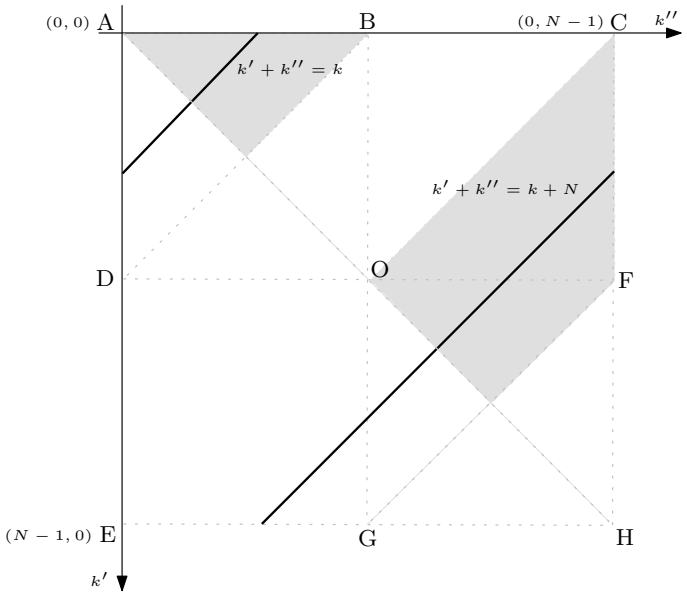

Fig. 3: Two-dimensional (k', k'') plane. Only points on the line $k'+k''=k$ (corresponding to sum interactions) and the line $k'+k''=k+N$ (corresponding to difference interactions) contribute to the output at the kth frequency bin.

values only at the upper triangle ACH. Therefore, the intersection between this triangle and the lines $k'+k''=k$ and $k'+k''=k+N$ bounds the range of the summation indices in (25), such that $d_{2;p,k}$ can be compactly rewritten as

$$d_{2;p,k} \approx \sum_{k' \in \mathcal{F}} x_{p,k'} x_{p,(k-k') \bmod N} c_{k',(k-k') \bmod N}\,, \tag{26}$$

where $\mathcal{F} = \{0, 1, \ldots \lfloor k/2 \rfloor, k+1, \ldots, k+1+\lfloor (N-k-2)/2 \rfloor\} \subset [0, N-1]$. Consequently, the number of cross-terms at the kth frequency bin has been reduced by a factor of two to $\lfloor k/2 \rfloor + \lfloor (N-k-2)/2 \rfloor + 2$. Note that a further reduction in the model complexity can be achieved if the signals are assumed real-valued, since in this case $c_{k',k''}$ must satisfy $c_{k',k''} = c^*_{N-k',N-k''}$, and thus, only points in the grey area contribute to the model output (in this case, it is sufficient to consider only the first $\lfloor N/2 \rfloor + 1$ output frequency bins).

It is worthwhile noting the aliasing effects in the model output signal. Aliasing exists in the output as a consequence of sum and difference interactions that produce frequencies higher than one-half of the Nyquist frequency. The input frequencies causing these aliasing effects correspond to the points in the triangles BDO and FGO. To avoid aliasing, one must require that the value of $x_{p,k'} x_{p,k''} c_{k',k''}$ is zero for all indices k' and k'' inside these triangles.

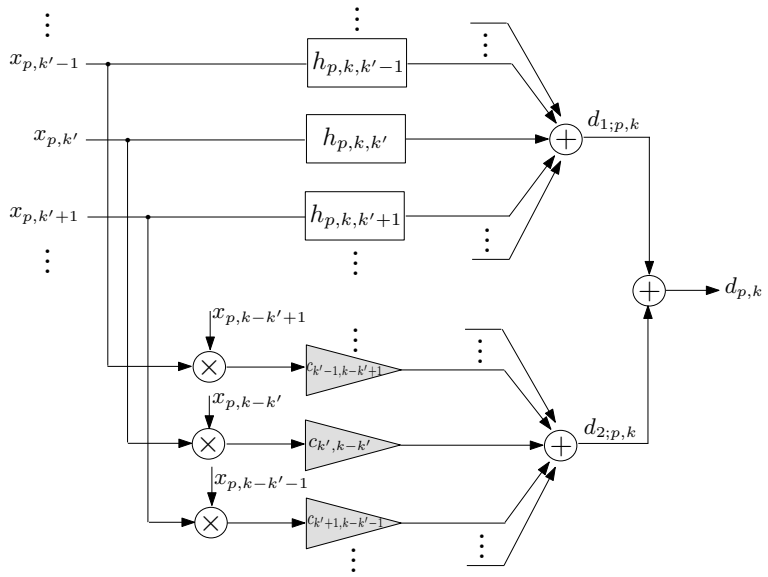

Fig. 4: Block diagram of the proposed model for quadratically nonlinear systems in the STFT domain. The upper branch represents the linear component of the system, which is modeled by the crossband filters $h_{p,k,k'}$. The quadratic component is modeled at the lower branch by using the quadratic cross-terms $c_{k,k'}$.

Finally, using (8) and (26) for representing the linear and quadratic components of the system, respectively, we obtain

$$d_{p,k} = \sum_{k'=0}^{N-1} \sum_{p'=0}^{\tilde{N}_1-1} x_{p-p',k'} h_{p',k,k'} + \sum_{k' \in \mathcal{F}} x_{p,k'} x_{p,(k-k') \bmod N} c_{k',(k-k') \bmod N} \,. \tag{27}$$

Equation (27) represents an explicit model for quadratically nonlinear systems in the STFT domain. A block diagram of the proposed model is illustrated in Fig. 4. Analogously to the time-domain Volterra model, an important property of the proposed model is the fact that its output depends linearly on the coefficients, which means that conventional linear estimation algorithms can be applied for estimating its parameters (see Section IV).

The proposed STFT-domain model generalizes the conventional discrete frequency-domain Volterra model [28], where the linear and quadratic components of the system are modeled in parallel using multiplicative terms:

$$D(k) = H_1(k)X(k) + \sum_{\substack{k',k''=0 \\ (k'+k'') \bmod N = k}}^{N-1} H_2(k',k'')X(k')X(k'') \,, \tag{28}$$

where $X(k)$ and $D(k)$ are the Nth-length discrete Fourier transforms (DFT's) of the input $x(n)$ and the output $d(n)$, respectively, and $H_1(k)$ and $H_2(k', k'')$ are the linear and quadratic Volterra transfer functions, respectively. A major limitation of this model is its underlying assumption that the observation frame (N) is sufficiently large compared with the memory length of the linear kernel, which enables to approximate the linear convolution as multiplicative in the frequency domain. Similarly, under this large-frame assumption, the linear component in the proposed model (27) can be approximated as a multiplicative transfer function (MTF) [39], [40]. Accordingly, the STFT model in (27) reduces to

$$d_{p,k} = h_k x_{p,k} + \sum_{k' \in \mathcal{F}} x_{p,k'} x_{p,(k-k') \bmod N} c_{k',(k-k') \bmod N}\,, \tag{29}$$

which is in one-to-one correspondence with the frequency-domain model (28). Therefore, the frequency-domain model can be regarded as a special case of the proposed model for relatively large observation frames. In practice, a large observation frame may be very restrictive, especially when long and time-varying impulse responses are considered (as in acoustic echo cancellation applications [41]). A long frame restricts the capability to identify and track time variations in the system, since the system is assumed constant during the observation frame. Additionally, as indicated in [39], increasing the frame length (while retaining the relative overlap between consecutive frames), reduces the number of available observations in each frequency bin, which increases the variance of the system estimate. Attempting to identify the system using the models (28) or (29) yields a model mismatch that degrades the accuracy of the linear-component estimate. The crossband filters representation, on the other hand, outperforms the MTF approach and achieves a substantially lower mse value, even when relatively long frames are considered [30]. Clearly, the proposed model forms a much richer representation than that offered by the frequency-domain model, and may correspondingly be useful for a larger variety of applications.

In this context, it should be emphasized that the quadratic-component representation provided by the proposed time-frequency model (27) (and certainly by the frequency-domain model) may not exactly represent a second-order Volterra filter in the time domain, due to the approximations made in (24) and (25). Nevertheless, the proposed STFT model forms a new class of nonlinear models that may represent certain nonlinear systems more efficiently than the conventional time-domain Volterra model. In fact, as will be shown in Section V, the proposed model may be more advantageous than the latter in representing nonlinear systems with relatively long memory due to its computational efficiency.

For completeness of discussion, let us extend the STFT model to the general case of a qth-order nonlinear system. Following a similar derivation to that made for the quadratic case [see (24)-(25)], the

output of a qth-order nonlinear system is modeled in the STFT domain as

$$d_{p,k} = d_{1;p,k} + \sum_{\ell=2}^{q} d_{\ell;p,k}\,, \tag{30}$$

where the linear component $d_{1;p,k}$ is given by (8), and the ℓth-order homogeneous component $d_{\ell;p,k}$ is given by

$$d_{\ell;p,k} = \sum_{\substack{k_{1,\ldots}k_\ell=0 \\ \left(\sum_{i=1}^{\ell} k_i\right) \bmod N = k}}^{N-1} c_{k_1,\ldots k_\ell} \prod_{i=1}^{\ell} x_{p,k_i}\,. \tag{31}$$

Clearly, only ℓ-fold frequencies $\{k_i\}_{i=1}^{\ell}$, whose sum is k or $k+N$, contribute to the output $d_{\ell;p,k}$ at frequency bin k. Consequently, the number of cross-terms $c_{k_1,\ldots k_{\ell-1},k_\ell}$ ($\ell = 2,\ldots,q$) involved in representing a qth-order nonlinear system is given by $\sum_{\ell=2}^{q} N^{\ell-1} = \left(N^q - N\right)/\left(N-1\right)$. Note that this number can be further reduced by exploiting the symmetry property of the cross-terms, as was done for the quadratic case.

IV. Quadratically Nonlinear System Identification

In this section, we consider the problem of identifying quadratically nonlinear systems using the proposed STFT model, and formulate an LS optimization criterion for estimating the model parameters in each frequency bin. The conventional time-domain Volterra filter identification is also described, and a comparison between the STFT- and time-domain models is carried out in terms of computational complexity. Without loss of generality, we consider here only the quadratic model due to its relatively simpler structure. The quadratic model is appropriate for representing the nonlinear behavior of many real world systems [42]. An extension to higher nonlinearity orders is straightforward.

Let an input $x(n)$ and output $y(n)$ of an unknown (quadratically) nonlinear system be related by

$$y(n) = \{\phi x\}(n) + \xi(n) = d(n) + \xi(n)\,, \tag{32}$$

where $\phi(\cdot)$ denotes a discrete-time nonlinear time-invariant system, $\xi(n)$ is a corrupting additive noise signal, and $d(n)$ is the clean output signal. Note that the "noise" signal $\xi(n)$ may sometimes include a useful signal, e.g., the local speaker signal in acoustic echo cancellation. The problem of system identification can be formulated as follows: Given an input signal $x(n)$ and noisy observation $y(n)$, construct a model for describing the input-output relationship, and select its parameters so that the model output $\hat{y}(n)$ best estimates (or predicts) the measured output signal. We denote by N_x the time-domain observable data length, and by $P \approx N_x/L$ the number of samples in a time-trajectory of the STFT representation (i.e., length of $x_{p,k}$ for a given k).

A. Identification in the STFT domain

A system identifier operating in the STFT domain is illustrated in Fig. 1. In the time-frequency domain, equation (32) may be written as

$$y_{p,k} = d_{p,k} + \xi_{p,k}\,. \tag{33}$$

To derive an estimator $\hat{y}_{p,k}$ for the system output in the STFT domain, we employ the quadratic STFT model proposed in the previous section [see (27)]. Utilizing only $2K$ crossband filters around each frequency bin for the estimation of the linear component, the resulting estimate $\hat{y}_{p,k}$ can be written as

$$\begin{aligned}\hat{y}_{p,k} \;=\; & \sum_{k'=k-K}^{k+K} \sum_{p'=0}^{\bar{N}_1-1} x_{p-p',k' \bmod N} h_{p',k,k' \bmod N} \\ & + \sum_{k' \in \mathcal{F}} x_{p,k'} x_{p,(k-k') \bmod N} c_{k',(k-k') \bmod N}\,. \end{aligned} \tag{34}$$

The influence of the number of estimated crossband filters ($2K+1$) on the system identifier performance is demonstrated in Section V.

Let $\mathbf{h}_k$ be the $2K+1$ filters at frequency bin k

$$\mathbf{h}_k = \left[\begin{array}{ccccc} \mathbf{h}_{k,(k-K) \bmod N}^T & \mathbf{h}_{k,(k-K+1) \bmod N}^T & \cdots & \cdots & \mathbf{h}_{k,(k+K) \bmod N}^T \end{array} \right]^T, \tag{35}$$

where $\mathbf{h}_{k,k'} = \left[\begin{array}{cccc} h_{0,k,k'} & h_{1,k,k'} & \cdots & h_{\bar{N}_1-1,k,k'} \end{array} \right]^T$ is the crossband filter from frequency bin k' to frequency bin k. Let $\mathbf{X}_k$ denote an $P \times M$ Toeplitz matrix whose (m,ℓ)th term is given by $(\mathbf{X}_k)_{m,\ell} = x_{m-\ell,k}$, and let $\mathbf{\Delta}_k$ be a concatenation of $\{\mathbf{X}_{k'}\}_{k'=(k-K) \bmod N}^{(k+K) \bmod N}$ along the column dimension

$$\mathbf{\Delta}_k = \left[\begin{array}{ccccc} \mathbf{X}_{(k-K) \bmod N} & \mathbf{X}_{(k-K+1) \bmod N} & \cdots & \cdots & \mathbf{X}_{(k+K) \bmod N} \end{array} \right]. \tag{36}$$

For notational simplicity, let us assume that k and N are both even, such that according to (26), the number of quadratic cross-terms in each frequency bin is $N/2+1$. Then, let

$$\mathbf{c}_k = \left[\begin{array}{ccccccc} c_{0,k} & \cdots & c_{\frac{k}{2},\frac{k}{2}} & c_{k+1,N-1} & \cdots & c_{\frac{N+k}{2},\frac{N+k}{2}} \end{array} \right]^T \tag{37}$$

denote the quadratic cross-terms at the kth frequency bin, and let

$$\mathbf{\Lambda}_k = \left[\begin{array}{ccccccc} \mathbf{x}_{0,k} & \cdots & \mathbf{x}_{\frac{k}{2},\frac{k}{2}} & \mathbf{x}_{k+1,N-1} & \cdots & \mathbf{x}_{\frac{N+k}{2},\frac{N+k}{2}} \end{array} \right] \tag{38}$$

be an $P \times (N/2+1)$ matrix, where $\mathbf{x}_{k,k'} = \left[\begin{array}{cccc} x_{0,k}x_{0,k'} & x_{1,k}x_{1,k'} & \cdots & x_{P-1,k}x_{P-1,k'} \end{array} \right]^T$ is a term-by-term multiplication of the time-trajectories of $x_{p,k}$ at frequency bins k and k', respectively. Then, the output signal estimate (34) can be written in a vector form as

$$\begin{aligned}\hat{\mathbf{y}}_k &= \mathbf{\Delta}_k \mathbf{h}_k + \mathbf{\Lambda}_k \mathbf{c}_k \\ &\triangleq \mathbf{R}_k \boldsymbol{\theta}_k\,, \end{aligned} \tag{39}$$

where $\hat{\mathbf{y}}_k = \left[\begin{array}{cccc} \hat{y}_{0,k} & \hat{y}_{1,k} & \cdots & \hat{y}_{P-1,k} \end{array}\right]^T$, $\mathbf{R}_k = [\boldsymbol{\Delta}_k \;\; \boldsymbol{\Lambda}_k]$, and $\boldsymbol{\theta}_k = \left[\mathbf{h}_k^T \;\; \mathbf{c}_k^T\right]^T$ is the model parameter vector. The dimension of $\boldsymbol{\theta}_k$ is given by

$$d_{\boldsymbol{\theta}_k} \triangleq \dim \boldsymbol{\theta}_k = (2K+1)\,\bar{N}_1 + N/2 + 1\,. \tag{40}$$

Denoting the observable data vector by $\mathbf{y}_k = \left[\begin{array}{cccc} y_{0,k} & y_{1,k} & \cdots & y_{P-1,k} \end{array}\right]^T$, and using the above notations, the LS estimate of the model parameters at the kth frequency bin is given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_k &= \arg\min_{\boldsymbol{\theta}_k} \|\mathbf{y}_k - \mathbf{R}_k\boldsymbol{\theta}_k\|^2 \\ &= \left(\mathbf{R}_k^H\mathbf{R}_k\right)^{-1}\mathbf{R}_k^H\mathbf{y}_k\,, \end{aligned} \tag{41}$$

where we assume that $\mathbf{R}_k^H\mathbf{R}_k$ is not singular[2]. Note that both $\hat{\boldsymbol{\theta}}_k$ and $\hat{\mathbf{y}}_k$ depend on the parameter K, but for notational simplicity K has been omitted. Substituting (41) into (39), we obtain an estimate of the system output in the STFT domain at the kth frequency bin. Repeating this estimation process for each frequency bin and returning to the time-domain using the inverse STFT (4), we obtain the system output estimator $\hat{y}_{\mathrm{s}}(n)$. The subscript s is to distinguish the subband-approach estimate from the fullband-approach estimate $\hat{y}_{\mathrm{f}}(n)$ [derived in Section IV-B].

Next, we evaluate the computational complexity of the proposed approach. Computing the parameter vector estimate $\hat{\boldsymbol{\theta}}_k$ requires a solution of the LS normal equations $\left(\mathbf{R}_k^H\mathbf{R}_k\right)\hat{\boldsymbol{\theta}}_k = \mathbf{R}_k^H\mathbf{y}_k$ for each frequency bin. This results in $Pd_{\hat{\boldsymbol{\theta}}_k}^2 + d_{\hat{\boldsymbol{\theta}}_k}^3/3$ arithmetic operations when using the Cholesky decomposition [44], where $d_{\boldsymbol{\theta}_k}$ is defined in (40). Computation of the desired signal estimate (39) requires additional $2Pd_{\boldsymbol{\theta}_k}$ arithmetic operations. Assuming P is sufficiently large, the complexity associated with the proposed model is

$$O_{\mathrm{s}} \sim O\left\{NP\left[(2K+1)\,\bar{N}_1 + N/2 + 1\right]^2\right\}. \tag{42}$$

Expectedly, we observe that the computational complexity increases as K increases. However, analogously to linear system identification [30], incorporating crossband filters into the model may yield lower mse for stronger and longer input signals, as demonstrated in Section V.

[2]In the ill-conditioned case, when $\mathbf{R}_k^H\mathbf{R}_k$ is singular, matrix regularization is required [43].

B. Identification in the time domain

For time-domain system identification, we utilize the second-order Volterra model, described in (1). Accordingly, an estimator for the system output can be expressed as

$$\begin{aligned}\hat{y}_{\mathrm{f}}(n) &= \sum_{m=0}^{N_1-1} h_1(m)x(n-m) \\ &\quad + \sum_{m=0}^{N_2-1}\sum_{\ell=m}^{N_2-1} h_2(m,\ell)x(n-m)x(n-\ell)\,, \end{aligned} \tag{43}$$

where for the quadratic kernel, the triangular Volterra representation is used [11].

Let $\mathbf{h}_1 = \left[\begin{array}{cccc} h_1(0) & h_1(1) & \cdots & h_1(N_1-1)\end{array}\right]^T$ denote the linear kernel, and let $\mathbf{x}_1(n) = \left[\begin{array}{cccc} x(n) & x(n-1) & \cdots & x(n-N_1+1)\end{array}\right]^T$. The quadratic kernel can be written in a vector notation as

$$\begin{aligned}\mathbf{h}_2 = \Big[& h_2(0,0) \quad h_2(0,1) \quad \cdots \quad h_2(0,N_2-1) \\ & h_2(1,1) \quad h_2(1,2) \quad \cdots \quad h_2(1,N_2-1) \\ & \cdots \quad h_2(N_2-1,N_2-1) \Big]^T . \end{aligned} \tag{44}$$

where similarly we define

$$\begin{aligned}\mathbf{x}_2(n) = \Big[& x^2(n) \quad x(n)\,x(n-1) \quad \cdots \quad x(n)\,x(n-N_2+1) \\ & x(n-1)\,x(n-1) \quad \cdots \quad x(n-1)\,x(n-N_2+1) \\ & \cdots \quad x^2(n-N_2+1) \Big]^T . \end{aligned} \tag{45}$$

Then, the system output estimate (43) can be written in a vector form as

$$\hat{y}_{\mathrm{f}}(n) = \mathbf{x}^T(n)\boldsymbol{\theta}\,, \tag{46}$$

where $\mathbf{x}(n) = \left[\mathbf{x}_1^T(n)\ \mathbf{x}_2^T(n)\right]$ and $\boldsymbol{\theta} \triangleq \left[\mathbf{h}_1^T\ \mathbf{h}_2^T\right]^T$ is the model parameter vector. Note that the dimension of $\boldsymbol{\theta}$, which determines the model complexity, is

$$d_{\boldsymbol{\theta}} \triangleq \dim \boldsymbol{\theta} = N_1 + \frac{N_2(N_2+1)}{2}\,. \tag{47}$$

Let $\mathbf{y} = \left[\begin{array}{cccc} y(0) & y(1) & \cdots & y(N_x-1)\end{array}\right]^T$, and let $\mathbf{X}$ be an $N_x \times d_{\boldsymbol{\theta}}$ matrix defined as $\mathbf{X}^T = \left[\begin{array}{cccc} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N_x-1)\end{array}\right]$. Then, the LS estimate of $\boldsymbol{\theta}$ is given by

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}} \|\mathbf{y} - \boldsymbol{X}\boldsymbol{\theta}\|^2 \\ &= \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{y}\,. \end{aligned} \tag{48}$$

Substituting (48) into (46), we obtain an estimate of the system output in the time domain $\hat{y}_{\mathrm{f}}(n)$ using a second-order Volterra model.

As in the subband approach, forming the normal equations, solving them using the Cholesky decomposition and calculating the desired signal estimate, require $N_x d_{\boldsymbol{\theta}}^2 + d_{\boldsymbol{\theta}}^3/3 + 2N_x d_{\boldsymbol{\theta}}$ arithmetic operations. For sufficiently large N_x, the computational complexity of the fullband approach can be expressed as

$$O_{\mathrm{f}} \sim O\left(N_x \left[N_1 + \frac{N_2\left(N_2+1\right)}{2}\right]^2\right). \tag{49}$$

It is worth noting that the complexity of the fullband approach can be generally reduced by using efficient algorithms that exploit the special structure of the corresponding matrix in the LS normal equations [45], [46].

C. Comparison and Discussion

Let $r = L/N$ denote the relative overlap between consecutive analysis windows (this overlap determines the redundancy of the STFT representation). Then, rewriting the subband approach complexity (42) in terms of the fullband parameters (by using the relations $P \approx N_x/L$ and $\bar{N}_1 \approx N_1/L$), the ratio between the fullband and subband complexities can be written as

$$\frac{O_{\mathrm{s}}}{O_{\mathrm{f}}} \sim \frac{1}{r} \cdot \frac{\left(2N_1 \cdot \frac{2K+1}{rN} + N\right)^2}{\left(2N_1 + N_2^2\right)^2}. \tag{50}$$

Expectedly, we observe that the computational gain achieved by the proposed subband approach is mainly determined by the STFT analysis window length N, which represents the trade-off between the linear- and nonlinear-component complexities. Specifically, using a longer analysis window yields shorter crossband filters ($\sim N_1/N$), which reduces the computational cost of the linear component, but at the same time increases the nonlinear-component complexity by increasing the number of quadratic cross-terms ($\sim N$). Nonetheless, according to (50), the complexity of the proposed subband approach would typically be lower than that of the conventional fullband approach. For instance, for $N = 256$, $r = 0.5$ (i.e., $L = 128$), $N_1 = 1024$, $N_2 = 80$ and $K = 2$ the proposed approach complexity is reduced by approximately 300, when compared to the fullband-approach complexity. The computational efficiency obtained by the proposed approach becomes even more significant when systems with relatively large second-order memory length are considered. This is because these systems necessitate an extremely large memory length N_2 for the quadratic kernel, when using the time-domain Volterra model, such that $N \ll N_2^2$ and consequently $O_{\mathrm{s}} \ll O_{\mathrm{f}}$.

An example of a long-memory system is an LEM system in nonlinear acoustic echo cancellation applications [3]–[5]. The nonlinear behavior of this system is mainly introduced by the loudspeakers and their amplifiers, especially when small loudspeakers are driven at high volume. When parallel models are considered for modeling the LEM system, the memory length of the nonlinear component will also be determined by the acoustic enclosure, which typically consists of several thousands taps [41]. Consequently, attempting to estimate the LEM system with the time-domain Volterra model involves high computational cost, which makes it impractical in real applications. To reduce the model complexity, the Volterra filters can be truncated in time [18], but then the system estimate is less accurate. Other time-domain approximations for Volterra filters employed for acoustic echo cancellation, such as the Hammerstein model (i.e., a static nonlinearity followed by a dynamic linear block, as in [3], [5]), suggest a less general structure than the Volterra filter. On the other hand, the proposed STFT model offers both structural generality and computational efficiency, which facilitate a practical alternative for the time-domain Volterra approach, especially in representing systems with long memory.

V. Experimental Results

In this section, we present experimental results that demonstrate the effectiveness of the proposed subband approach in estimating and modeling quadratically nonlinear systems. A comparison to the conventional time-domain Volterra approach is carried out in terms of mse performance for both synthetic white Gaussian signals and real speech signals. The evaluation includes objective quality measures, a subjective study of temporal waveforms, and informal listening tests. For the STFT, we use half overlapping Hamming analysis windows of $N = 256$ samples length (i.e., $L = 0.5N$). The inverse STFT is implemented with a minimum-energy synthesis window that satisfies the completeness condition [36].

A. Performance Evaluation for White Gaussian Input Signals

In the first experiment, we examine the performances of the Volterra and proposed models under the assumption of white Gaussian signals. The system to be identified is formed as a parallel combination of linear and quadratic components as follows:

$$y(n) = \sum_{m=0}^{N_1^*-1} g_1(m)x(n-m) + \{\mathcal{L}x\}(n) + \xi(n), \tag{51}$$

where $g_1(n)$ is the true linear kernel and $\{\mathcal{L}x\}(n)$ denotes the output of the quadratic component. The input signal $x(n)$ and the additive noise signal $\xi(n)$ are uncorrelated zero-mean white Gaussian processes with variances σ_x^2 and σ_ξ^2, respectively. We model the linear kernel as a nonstationary stochastic process

with an exponential decay envelope, i.e., $g_1(n) = u(n)\beta(n)e^{-\alpha n}$, where $u(n)$ is the unit step function, $\beta(n)$ is a unit-variance zero-mean white Gaussian noise, and α is the decay exponent. In the following, we use $N_1^* = 768$, $\alpha = 0.009$, and an observable data length of $N_x = 24000$ samples. For evaluating the quality of the system estimate, the normalized mse is defined as

$$\epsilon_\gamma = \frac{E\left\{\left|d\left(n\right) - \hat{y}_\gamma\left(n\right)\right|^2\right\}}{E\left\{\left|d\left(n\right)\right|^2\right\}}, \tag{52}$$

where $d(n)$ is the clean output signal [i.e., $d(n) = y(n) - \xi(n)$], $\gamma \in \{\mathrm{s}, \mathrm{f}\}$, and $\hat{y}_\mathrm{s}(n)$ and $\hat{y}_\mathrm{f}(n)$ are the system output estimates obtained by the proposed subband approach and the fullband Volterra approach, respectively (see Section IV).

In the first experiment, we assume that the output signal of the true-system's quadratic component $\{\mathcal{L}x\}\left(n\right)$ is generated according to the quadratic model proposed in (26). That is, denoting by S^{-1} the inverse STFT operator, $\{\mathcal{L}x\}\left(n\right)$ can be expressed as

$$\{\mathcal{L}x\}\left(n\right) = S^{-1}\sum_{k'\in\mathcal{F}} x_{p,k'}x_{p,(k-k')\bmod N}g_{k',(k-k')\bmod N}\,, \tag{53}$$

where $\left\{\, g_{k',(k-k')\bmod N}\,\middle|\, k' \in \mathcal{F}\right\}$ are the true quadratic cross-terms. These terms are modeled here as a unit-variance zero-mean white Gaussian process. For both models, a memory length of $N_1 = 768$ is employed for the linear kernel, where the memory length N_2 of the quadratic kernel in the Volterra model is set to 30. Figure 5 shows the resulting mse curves as a function of the SNR [the SNR is defined as the power ratio between the clean output signal $d(n)$ and the additive noise signal $\xi(n)$], as obtained for a nonlinear-to-linear ratio (NLR) of 0 dB [Fig. 5(a)] and -20 dB [Fig. 5(b)]. The NLR represents the power ratio between the output signals of the quadratic and linear components of the true system. For the proposed model, several values of K are employed in order to determine the influence of the number of estimated crossband filters on the mse performance, and the optimal value that achieves the minimal mse (mmse) is indicated above the mse curve. Note that a transition in the value of K is indicated by a variation in the width of the curve. Figure 5(a) implies that for relatively low SNR values, a lower mse is achieved by the conventional Volterra model. For instance, for an SNR of -20 dB, employing the Volterra model reduces the mse by approximately 10 dB, when compared to that achieved by the proposed model. However, for higher SNR conditions, the proposed model is considerably more advantageous. For an SNR of 20 dB, for instance, the proposed model enables a decrease of 17 dB in the mse using $K = 4$ (i.e., by incorporating 9 crossband filters into the model). Table I specifies the mse values obtained by each value of K for various SNR conditions. We observe that for high SNR values a significant improvement over the Volterra model can also be attained by using only the band-to-band

filters (i.e., $K = 0$), which further reduces the computational cost of the proposed model. Clearly, as the SNR increases, a larger number of crossband filters should be utilized to attain the mmse, which is similar to what has been shown in the identification of purely linear systems [30]. Note that similar results are obtained for a smaller NLR value [Fig. 5(b)], with the only difference is that the two curves intersect at a higher SNR value.

Figure 5 also provides an insight into the influence of undermodeling errors on the mse performance. Undermodeling errors occur whenever a given model does not admit an exact description of the true system. In our case, the undermodeling error of the Volterra model is due to the nonlinear component of the system, which cannot be accurately described by a second-order homogeneous Volterra filter. In the proposed model, on the other hand, the undermodeling error is a consequence of restricting the number of crossband filters in the linear component of the model [while the system's nonlinear component (53) can be perfectly represented by the model]. These undermodeling errors cause the mse curves of both models to saturate. The saturation values of the Volterra model and the proposed model, for any value of K (except for $K = 4$), are given at the right column of Table I (35 dB SNR). For the $K = 4$ mse curve, the saturation is attained at a relatively high SNR value [approximately 80 dB, for a 0 dB NLR; not displayed in Figure 5(a)]. This may be attributable to the fact that the linear component of the system can be represented almost perfectly with only four crossband filters around each frequency bin [30], such that the undermodeling error in this case becomes insignificant. Furthermore, a comparison of Figs. 5(a) and (b) indicates that the saturated mse value of the Volterra model decreases as the NLR decreases, which stems from the fact that the error induced by the undermodeling in the nonlinear component becomes less substantial as the nonlinearity strength decreases.

The complexity of the fullband and subband approaches (for each value of K) is evaluated by computing the central processing unit (CPU) running time[3] of the LS estimation process. The running time in terms of CPU seconds is averaged over several SNR conditions and summarized in Table II. We observe, as expected from (50), that the running time of the proposed approach, for any value of K, is substantially lower than that of the Volterra approach. Specifically, the estimation process of the Volterra model is approximately 12 and 4.5 times slower than that of the proposed model with $K = 0$ and $K = 4$, respectively. Moreover, Table II indicates that the running time of the proposed approach increases as more crossband filters are estimated, as expected from (42).

[3]The simulations were all performed under MATLAB; v.7.0, on a Core(TM)2 Duo P8400 2.27 GHz PC with 4 GB of RAM, running Windows Vista, Service Pack 1.

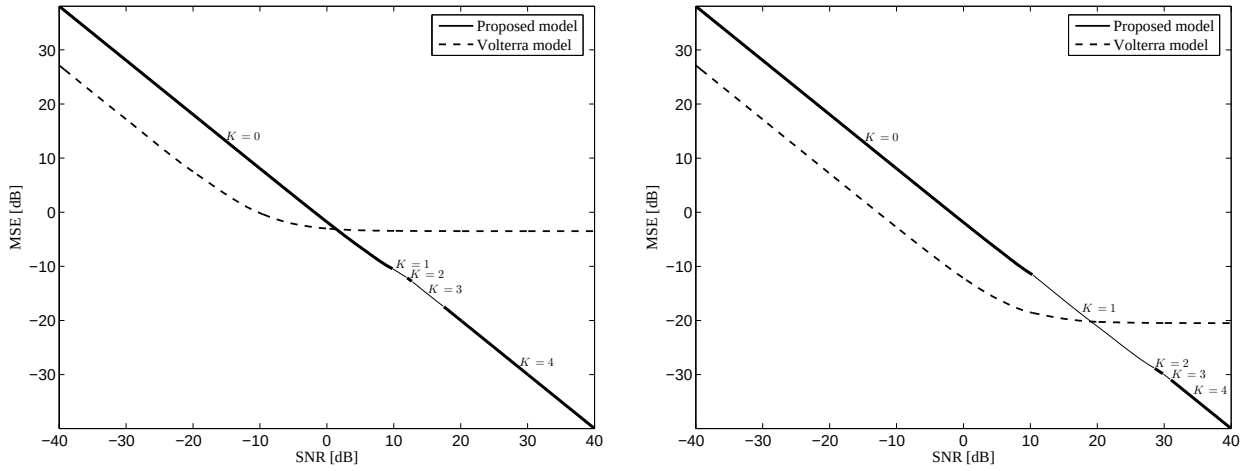

Fig. 5: MSE curves as a function of the SNR for white Gaussian signals, as obtained by the proposed STFT model (34) and the conventional time-domain Volterra model (43). The optimal value of K is indicated above the corresponding mse curve. The true system is formed as a combination of linear and quadratic components, where the latter is modeled according to (53). (a) Nonlinear-to-linear ratio (NLR) of 0 dB (b) NLR of -20 dB.

TABLE I: MSE Obtained by the Proposed Model for Several K Values and by the Volterra Model, Under Various SNR Conditions. The Nonlinear-to-Linear Ratio (NLR) is 0 dB.

K	MSE [dB]		
	SNR$= -10$ dB	SNR$= 20$ dB	SNR$= 35$ dB
0	8.08	-15.12	-16.05
1	8.75	-16.91	-18.8
2	9.31	-18.17	-21.55
3	9.82	-19.67	-28.67
4	10.04	-19.97	-34.97
Volterra	0.42	-3.25	-3.58

Next, we compare the Volterra and proposed models for a quadratically nonlinear system with a relatively large memory length. We assume that the quadratic component of the true system $\{\mathcal{L}x\}(n)$ is given by

$$\{\mathcal{L}x\}(n) = \sum_{m=0}^{N_1^*-1} g_1(m)x^2(n-m)\,, \tag{54}$$

where $g_1(n)$ is similar to that used in the previous experiment. A system represented by (51) and (54)

TABLE II: Average Running Time in Terms of CPU of the Proposed Approach (for Several K Values) and the Volterra Approach. The Length of the Observable Data is 24000 Samples.

K	Running Time [sec]
0	5.15
1	6.79
2	8.64
3	10.78
4	13.23
Volterra	61.31

can be viewed as a memoryless polynomial of the form $x(n)+x^2(n)$ followed by the linear kernel $g_1(n)$. Such a representation has been employed in acoustic echo cancellation applications, where memoryless nonlinearities occur in the power amplifier of the loudspeaker [5], [23]. Note that the memory length of the quadratic component is now equal to that of the linear component, and therefore, large values of N_2 should be used in the Volterra model in order to achieve satisfactory results. Figure 6 shows the resulting mse curves as a function of the SNR, where for the Volterra model, a relatively small memory length ($N_2 = 40$) and a large one ($N_2 = 80$) are used. Clearly, as the SNR increases, the proposed model outperforms the Volterra model (even for long kernels) and yields the mmse. For instance, for an SNR of 25 dB, an improvement of 16 dB can be achieved by using the proposed model rather than the Volterra model with $N_2 = 80$.

We observe that as the SNR increases, the mse performance of the Volterra model can be generally improved by using a longer memory for the quadratic kernel [at the expense of a considerable increase in computational complexity, as indicated by (49)]. This phenomenon is related to the problem of model-order selection, a fundamental problem in many system identification applications [1], [47]–[52], where in our case the model order is determined by the memory length of the quadratic Volterra kernel. Generally, the optimal model order is affected by the level of noise in the data and the length of the observable data. As the SNR increases or as more data is employable, the optimal model complexity increases, and correspondingly longer quadratic kernels can be utilized to achieve lower mse. The same reasoning is also relevant to explaining why the number of estimated crossband filter in the proposed subband model increases for larger SNRs. The experimental results show that a Volterra model in the time domain is not sufficient for identification of nonlinear systems with relatively long memory. The advantage of the proposed model is demonstrated in estimation accuracy and computational efficiency.

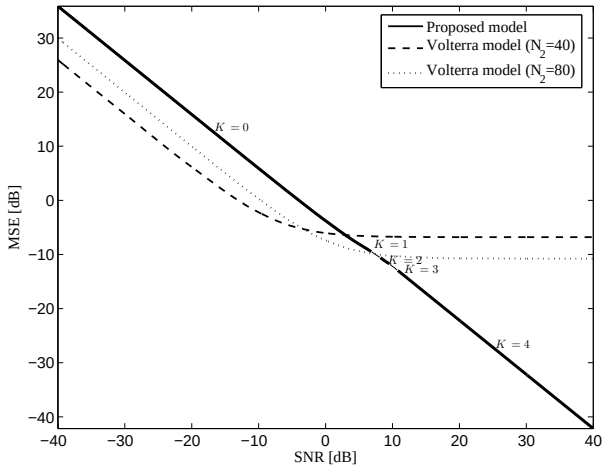

Fig. 6: MSE curves as a function of the SNR for white Gaussian signals, as obtained by the proposed STFT model (34) and the conventional time-domain Volterra model (43). The true system is formed as a memoryless polynomial of the form $x(n) + x^2(n)$ followed by a linear block.

B. Acoustic Echo Cancellation Scenario

In the second experiment, we demonstrate the application of the proposed approach to acoustic echo cancellation using real speech signals. We use an ordinary office with a reverberation time T_{60} of about 100 ms. A far-end speech signal $x(n)$ is fed into a loudspeaker at high volume, thus introducing non-negligible nonlinear distortion. The signal $x(n)$ propagates through the enclosure and received by a microphone as an echo signal together with a local noise $\xi(n)$. The resulting noisy signal is denoted by $y(n)$. In this experiment, the signals are sampled at 16 kHz. Note that the acoustic echo canceller (AEC) performance is evaluated in the absence of near-end speech, since a double-talk detector (DTD) is usually employed for detecting the near-end signal and freezing the estimation process [53], [54]. A commonly-used quality measure for evaluating the performance of AECs is the echo-return loss enhancement (ERLE), defined in dB by

$$\text{ERLE}_\gamma = 10 \log_{10} \frac{E\left\{y^2(n)\right\}}{E\left\{e_\gamma^2(n)\right\}}, \tag{55}$$

where

$$e_\gamma(n) = y(n) - \hat{y}_\gamma(n) \tag{56}$$

is the error signal (or residual echo signal) and $\hat{y}_\gamma(n)$ is defined in (52).

Figures 7(a) and (b) show the far-end signal and the microphone signal, respectively. Figures 7(c)–(e) show the error signals as obtained by using a purely linear model in the time domain, a Volterra model

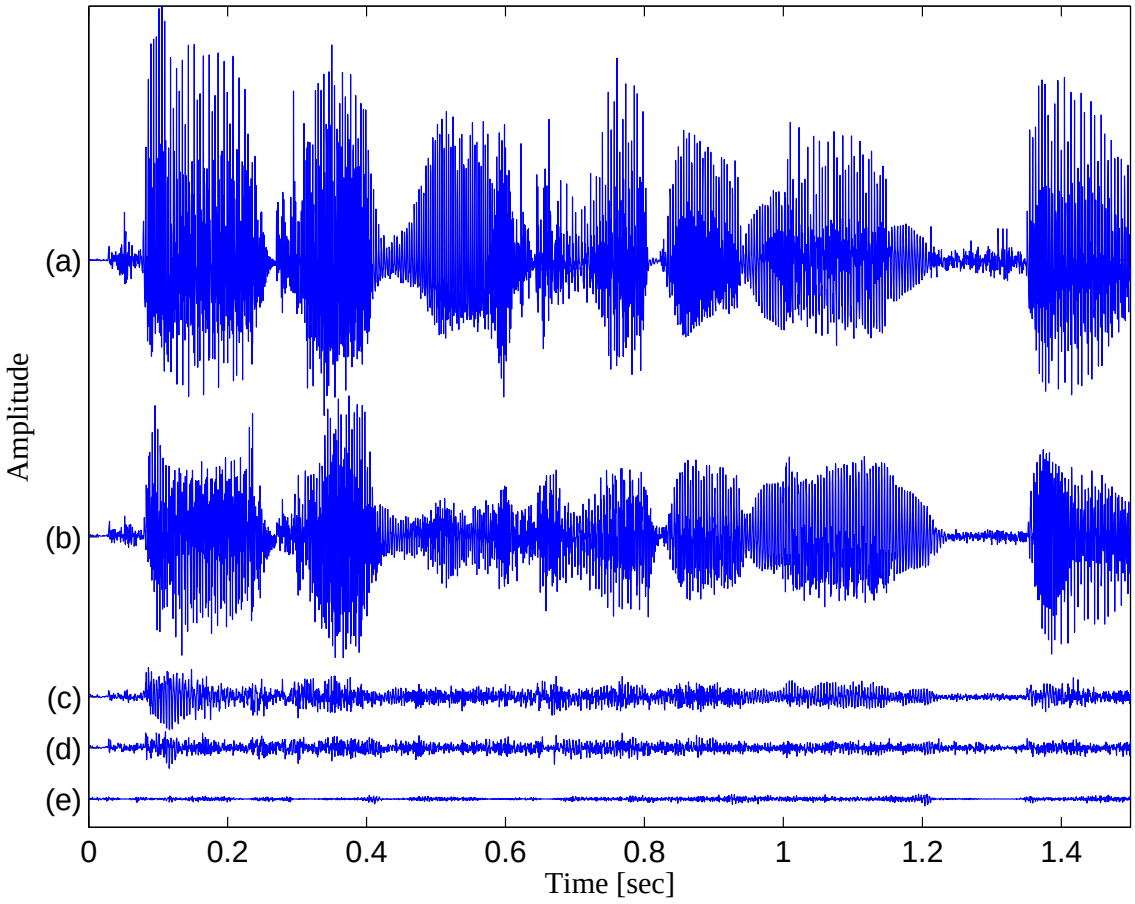

Fig. 7: Speech waveforms and residual echo signals, obtained by the time-domain Volterra approach and the proposed subband approach. (a) Far-end signal (b) Microphone signal. (c)–(e) Error signals obtained by a purely linear model in the time domain, the Volterra model with $N_2 = 90$, and the proposed model with $K = 1$, respectively. For all models, a length of $N_1 = 768$ is assumed in the linear kernel.

with $N_2 = 90$, and the proposed model with $K = 1$, respectively. For all models, a length of $N_1 = 768$ is employed for the linear kernel. The ERLE values of the corresponding error signals were computed by (55), and are given by 14.56 dB (linear), 19.14 dB (Volterra), and 29.54 dB (proposed). Clearly, the proposed approach achieves a significant improvement over a time domain approach. This may be attributable to the long memory of the system's nonlinear components which necessitate long kernels for sufficient modeling of the acoustic path. Furthermore, a purely linear model does not provide a sufficient echo attenuation due to nonlinear undermodeling [55]–[57]. Subjective listening tests confirm that the proposed approach achieves a perceptual improvement in speech quality over the conventional Volterra approach (audio files are available on-line [58]).

VI. CONCLUSIONS

Motivated by the common drawbacks of conventional time- and frequency-domain methods, we have introduced a novel approach for identifying nonlinear systems in the STFT domain. We have derived an explicit nonlinear model, based on an efficient approximation of Volterra-filters representation in the time-frequency domain. The proposed model consists of a parallel combination of a linear component, which is represented by crossband filters between subbands, and a nonlinear component, modeled by multiplicative cross-terms. We showed that the conventional discrete frequency-domain model is a special case of the proposed model for relatively long observation frames. Furthermore, we considered the identification of quadratically nonlinear systems and showed that a significant reduction in computational cost can be achieved over the time-domain Volterra model by the proposed approach. Experimental results have demonstrated the advantage of the proposed STFT model in estimating nonlinear systems with relatively large memory length. The time-domain Volterra model fails to estimate such systems due to its high complexity. The proposed model, on the other hand, achieves a significant improvement in mse performance, particularly for high SNR conditions. It is worthwhile noting, though, that the experimental results presented in this paper are applicable only for purely quadratic systems. When higher nonlinearity orders are considered, one should employ the extended STFT model [see (30)-(31)] and follow a similar identification process to that made for the quadratic case.

Overall, the results have met the expectations originally put into STFT-based estimation techniques. The proposed approach in the STFT domain offers both structural generality and computational efficiency, and consequently facilitates a practical alternative for conventional methods.

Since practically many real-world systems are time-varying, the approach proposed in this paper should be made adaptive in order to track these variations. Recently, an adaptive estimation of the model parameters and a detailed convergence analysis of the adaptation process was introduced [59]. Future research will concentrate on constructing a fully adaptive-control scheme, which exploits the attractive properties of the proposed model and provides a balance between complexity, convergence rate and steady-state performance.

APPENDIX A
DERIVATION OF (11)

Using (2) and (1), the STFT of $d_2(n)$ can be written as

$$d_{2;p,k} = \sum_{n,m,\ell} h_2(m,\ell)x(n-m)x(n-\ell)\ \tilde{\psi}^*_{p,k}(n) \tag{57}$$

Substituting (4) into (57), we obtain

$$\begin{aligned} d_{2;p,k} &= \sum_{n,m,\ell} h_2(m,\ell) \sum_{k'=0}^{N-1} \sum_{p'} x_{p',k'} \psi_{p',k'}(n-m) \\ &\quad \times \sum_{k''=0}^{N-1} \sum_{p''} x_{p'',k''} \psi_{p'',k''}(n-\ell) \tilde{\psi}^*_{p,k}(n) \\ &= \sum_{k',k''=0}^{N-1} \sum_{p',p''} x_{p',k'} x_{p'',k''} c_{p,p',p'',k,k',k''} \end{aligned} \tag{58}$$

where

$$c_{p,p',p'',k,k',k''} = \sum_{n,m,\ell} h_2(m,\ell) \psi_{p',k'}(n-m) \psi_{p'',k''}(n-\ell) \tilde{\psi}^*_{p,k}(n) \,. \tag{59}$$

Substituting (3) and (5) into (59), we obtain

$$\begin{aligned} c_{p,p',p'',k,k',k''} &= \sum_{n,m,\ell} h_2(m,\ell) \psi(n-m-p'L) e^{j\frac{2\pi}{N}k'(n-m-p'L)} \\ &\quad \times \psi(n-\ell-p''L) e^{j\frac{2\pi}{N}k''(n-\ell-p''L)} \tilde{\psi}^*(n-pL) e^{-j\frac{2\pi}{N}k(n-pL)} \\ &= \sum_{n,m,\ell} h_2(m,\ell) \psi\left(\left(p-p'\right)L+n-m\right) e^{j\frac{2\pi}{N}k'((p-p')L+n-m)} \\ &\quad \times \psi\left(\left(p-p''\right)L+n-\ell\right) e^{j\frac{2\pi}{N}k''((p-p'')L+n-\ell)} \tilde{\psi}^*(n) e^{-j\frac{2\pi}{N}kn} \\ &= \{h_2(n,m) * \phi_{k,k',k''}(n,m)\}|_{n=(p-p')L,\ m=(p-p'')L} \triangleq c_{p-p',p-p'',k,k',k''} \end{aligned} \tag{60}$$

where $*$ denotes a 2D convolution with respect to the time indices n and m, and

$$\phi_{k,k',k''}(n,m) \triangleq \sum_{\ell} \tilde{\psi}^*(\ell) e^{-j\frac{2\pi}{N}k\ell} \psi(n+\ell) e^{j\frac{2\pi}{N}k'(n+\ell)} \psi(m+\ell) e^{j\frac{2\pi}{N}k''(m+\ell)} \,. \tag{61}$$

¿From (60), $c_{p,p',p'',k,k',k''}$ depends on $(p-p')$ and $(p-p'')$ rather than on p, p' and p'' separately. Substituting (60) into (58), we obtain (11).

ACKNOWLEDGEMENT

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

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999.

[2] R. Pintelon and J. Schoukens, *System Identification: A frequency domain approach*. Piscataway, NJ: IEEE Press, 2001.

[3] H. Dai and W. P. Zhu, "Compensation of loudspeaker nonlinearity in acoustic echo cancellation using raised-cosine function," *IEEE Trans. Circuits Syst. II*, vol. 53, no. 11, pp. 1190–2006, Nov. 2006.

[4] A. Guérin, G. Faucon, and R. L. Bouquin-Jeannès, "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 6, pp. 672–683, Nov. 2003.

[5] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, no. 9, pp. 1747–1760, 2000.

[6] S. Benedetto and E. Biglieri, "Nonlinear equalization of digital satellite channels," *IEEE J. Select. Areas Commun.*, vol. SAC-1, pp. 57–62, Jan. 1983.

[7] D. G. Lainiotis and P. Papaparaskeva, "A partitioned adaptive approach to nonlinear channel equalization," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1325–1336, Oct. 1998.

[8] D. T. Westwick and R. E. Kearney, "Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics," *Annals of Biomedical Engineering*, vol. 29, no. 8, pp. 707–718, Aug. 2001.

[9] G. Ramponi and G. L. Sicuranza, "Quadratic digital filters for image processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 6, pp. 937–939, Jun. 1988.

[10] F. Gao and W. M. Snelgrove, "Adaptive linearization of a loudpseaker," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada, May 1991, pp. 3589–3592.

[11] W. J. Rugh, *Nonlinear System Theory: The Volterra-Wiener Approach*. John Hopkins Univ. Press, 1981.

[12] T. Koh and E. J. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 6, pp. 1445–1455, Dec. 1985.

[13] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Krieger, 1989.

[14] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Processing Mag.*, vol. 8, no. 3, pp. 10–26, Jul. 1991.

[15] G. O. Glentis, P. Koukoulas, and N. Kalouptsidis, "Efficient algorithms for Volterra system identification," *IEEE Trans. Signal Processing*, vol. 47, no. 11, pp. 3042–3057, Nov. 1999.

[16] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York: Wiley, 2000.

[17] A. Fermo, A. Carini, and G. L. Sicuranza, "Simplified Volterra filters for acoustic echo cancellation in GSM receivers," in *European Signal Processing Conf.*, Tampere, Finland, 2000.

[18] A. Stenger., L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, USA, Mar. 1999, pp. 877–880.

[19] E. Biglieri, A. Gersho, R. D. Gitlin, and T. L. Lim, "Adaptive cancellation of nonlinear intersymbol interference for voiceband data transmission," *IEEE J. Select. Areas Commun.*, vol. 2, no. 5, pp. 765–777, Sep. 1984.

[20] R. D. Nowak, "Penalized least squares estimation of Volterra filters and higher order statistics," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 419–428, Feb. 1998.

[21] S. Im and E. J. Powers, "A block LMS algorithm for third-order frequency-domain Volterra filters," *IEEE Signal Processing Lett.*, vol. 4, no. 3, pp. 75–78, Mar. 1997.

[22] R. D. Nowak and B. D. V. Veen, "Random and pseudorandom inputs for Volterra filter identification," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2124–2135, Aug. 1994.

[23] F. Kuech and W. Kellermann, "Orthogonalized power filters for nonlinear acoustic echo cancellation," *Signal Processing*, vol. 86, pp. 1168–1181, 2006.

[24] E. W. Bai and M. Fu, "A blind approach to Hammerstein model identification," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1610–1619, Jul. 2002.

[25] T. M. Panicker, "Parallel-cascade realization and approximation of truncated Volterra systems," *IEEE Trans. Signal Processing*, vol. 46, no. 10, pp. 2829–2832, Oct. 1998.

[26] W. A. Frank, "An efficient approximation to the quadratic Volterra filter and its application in real-time loudspeaker linearization," *Signal Processing*, vol. 45, pp. 97–113, 1995.

[27] P. Koukoulas and N. Kalouptsidis, "Nonlinear system identification using Gaussian inputs," *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 1831–1841, Aug. 1995.

[28] K. I. Kim and E. J. Powers, "A digital method of modeling quadratically nonlinear systems with a general random input," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 11, pp. 1758–1769, Nov. 1988.

[29] C. H. Tseng and E. J. Powers, "Batch and adaptive Volterra filtering of cubically nonlinear systems with a Gaussian input," in *IEEE Int. Symp. Circuits and Systems (ISCAS)*, vol. 1, 1993, pp. 40–43.

[30] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio Speech Lang. Processing*, vol. 15, no. 4, pp. 1305–1319, May 2007.

[31] C. Faller and J. Chen, "Suppressing acoustic echo in a spectral envelope space," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 13, no. 5, pp. 1048–1062, Sep. 2005.

[32] Y. Lu and J. M. Morris, "Gabor expansion for adaptive echo cancellation," *IEEE Signal Processing Mag.*, vol. 16, pp. 68–80, Mar. 1999.

[33] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Processing*, vol. 40, no. 8, pp. 1862–1875, Aug. 1992.

[34] Y. Avargel and I. Cohen, "Adaptive system identification in the short-time Fourier transform domain using cross-multiplicative transfer function approximation," *IEEE Trans. Audio Speech Lang. Processing*, vol. 16, no. 1, pp. 162–173, Jan. 2008.

[35] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Signal Processing*, vol. ASSP-28, no. 1, pp. 55–69, Feb. 1980.

[36] J. Wexler and S. Raz, "Discrete Gabor expansions," *Signal Processing*, vol. 21, pp. 207–220, Nov. 1990.

[37] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Trans. Signal Processing*, vol. 41, no. 7, pp. 2429–2438, Jul. 1993.

[38] Y. Avargel and I. Cohen, "Identification of linear systems with adaptive control of the cross-multiplicative transfer function approximation," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, Las Vegas, Nevada, Apr. 2008, pp. 3789 – 3792.

[39] ——, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Processing Lett.*, vol. 14, no. 5, pp. 337–340, May 2007.

[40] ——, "Nonlinear acoustic echo cancellation based on a multiplicative transfer function approximation," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Seattle, WA, USA, Sep. 2008, pp. 1 – 4, paper no. 9035.

[41] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tlip, "Acoustic echo control," *IEEE Signal Processing Mag.*, vol. 16, no. 4, pp. 42–69, Jul. 1999.

[42] G. L. Sicuranza, "Quadratic filters for signal processing," *Proc. IEEE*, vol. 80, no. 8, pp. 1263–1285, Aug. 1992.

[43] A. Neumaier, "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM Rev.*, vol. 40, no. 3, pp. 636–666, Sep. 1998.

[44] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins University Press, 1996.

[45] G. Glentis and N. Kalouptsidis, "Efficient multichannel FIR filtering using a step versatile order recursive algorithm," *Signal Processing*, vol. 37, no. 3, pp. 437–462, Jun. 1994.

[46] ——, "Efficient order recursive algorithms for multichannel least squares filtering," *IEEE Trans. Signal Processing*, vol. 40, no. 6, pp. 1354–1374, Jun. 1992.

[47] F. D. Ridder, R. Pintelon, J. Schoukens, and D. P. Gillikin, "Modified AIC and MDL model selection criteria for short data records," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 1, pp. 144–150, Feb. 2005.

[48] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

[49] P. Stoica and Y. Selen, "Model order selection: a review of information criterion rules," *IEEE Signal Processing Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[50] G. C. Goodwin, M. Gevers, and B. Ninness, "Quantifying the error in estimated transfer functions with application to model order selection," *IEEE Trans. Autom. Control*, vol. 37, no. 7, pp. 913–928, Jul. 1992.

[51] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[52] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[53] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 2, pp. 168–172, Mar. 2000.

[54] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. Speech Audio Processing*, vol. 7, no. 6, pp. 718–724, Nov. 1999.

[55] A. E. Nordsjo, B. M. Ninness, and T. Wigren, "Quantifying model error caused by nonlinear undermodeling in linear system identification," in *Preprints 13th World Congr. IFAC*, vol. I, San Francisco, CA, 1996, pp. 145–149.

[56] B. Ninness and S. Gibson, "Quantifying the accuracy of hammerstein model estimation," *Automatica*, vol. 38, no. 12, pp. 2037–2051, 2002.

[57] J. Schoukens, R. Pintelon, T. Dobrowiecki, and Y. Rolain, "Identification of linear systems with nonlinear distortions," *Automatica*, vol. 41, no. 3, pp. 491–504, 2005.

[58] Y. Avargel Homepage. [Online]. Available: http://sipl.technion.ac.il/˜yekutiel

[59] Y. Avargel and I. Cohen, "Adaptive nonlinear system identification in the short-time Fourier transform domain," *to appear in IEEE Trans. Signal Processing*.