# A Comparison of the Squared Energy and Teager-Kaiser Operators for Short-Term Energy Estimation in Additive Noise

Dimitrios Dimitriadis *Member, IEEE*, Alexandros Potamianos *Member, IEEE*, and Petros Maragos *Fellow, IEEE*

**Abstract**

Time-frequency distributions that evaluate the signal's energy content both in the time and frequency domains are indispensable signal processing tools, especially, for non-stationary signals. Various short-time energy computation schemes are used in practice, including the mean squared amplitude and Teager-Kaiser energy approaches. Herein, we focus primarily on the short- and medium-term properties of these two energy estimation schemes, as well as, on their performance in the presence of additive noise. To facilitate this analysis and generalize the approach, we use a harmonic noise model to approximate the noise component. The error analysis is conducted both in the continuous- and discrete-time domains, deriving similar conclusions. The estimation errors are measured in terms of normalized deviations from the expected signal energy and are shown to greatly depend on both the signals' spectral content and the analysis window length. When medium- and long-term analysis windows are employed, the Teager-Kaiser energy operator is proven superior to the common squared energy operator, provided that the spectral content of the noise is more lowpass than the corresponding signal content, and vice versa.



D. Dimitriadis and P. Maragos are with the School of Electrical & Computer Engineering, National Technical University of Athens, Zografou, Athens, GR-15773, Greece; email: {ddim, maragos}@cs.ntua.gr; tel: +30 210 7722455; fax: +30 210 7723397.

A. Potamianos is with the Dept. of Electronics & Computer Engineering, Technical Univ. of Crete, Chania, GR-73100, Greece; email: potam@telecom.tuc.gr.

Manuscript received Aug. 05, 2008; revised Dec. 23, 2008; accepted Feb. 04, 2009. This work was supported in part by the European FP6-IST Network of Excellence 'MUSCLE' (IST-FP6-507752), and in part by the project ΠΕΝΕΔ-2003 ΕΔ-866, which is co-financed by the E.U.-European Social Fund (80%) and the Greek Ministry of Development-GSRT (20%).

However, for shorter window lengths, the Teager-Kaiser operator always outperforms the squared energy operator. The theoretical results are experimentally verified for synthetic signals. Finally, the performance of the proposed energy operators is evaluated for short-term analysis of noisy speech signals and the implications for speech processing applications are outlined.

**Index Terms**

Time-frequency analysis, robustness, harmonic analysis, noise, spectral analysis, bandlimited signals, feature extraction, signal detection, estimation.

**EDICS Category: NSP-APPL: Applications of nonlinear signal processing, DSP-TFSR: Time-frequency analysis and signal representation**

## I. INTRODUCTION

Time-frequency distributions estimating the signal energy content in time and frequency bins are considered indispensable for the study of non-stationary signals. Such signals frequently appear in many applications, including speech, radar, geophysical, biological, and transient signal analysis and processing. In this context, various time-frequency distributions have been studied and implemented [5], [9], with some generalizations found in [1].

In signal processing applications, signals are often corrupted by noise, attributed to the environment, sensor or channel. Thus, the computation of such time-frequency distributions can be generalized as an energy estimation problem in the presence of noise. Robust energy estimation is a complex problem, much studied over the years. Despite these intensive research efforts, certain aspects still remain under-researched. Moreover, the extension of these ideas to the discrete-time domain is neither clear nor straightforward. The most widely used energy estimation scheme is based on the *Squared Energy Operator (SEO)* $S[\cdot]$, where the squared signal is the desired instantaneous energy term [25]:

$$S\left[x(t)\right] \triangleq x^2(t) \tag{1}$$

An alternative scheme is based on the *Teager-Kaiser Energy Operator (TEO)* [15], [20], [21]

$$\Psi\left[x(t)\right] \triangleq \dot{x}^2(t) - x(t)\ddot{x}(t) \tag{2}$$

where $\dot{x}(t) = dx(t)/dt$. This latter nonlinear operator approach has been mainly used for the energy estimation of AM-FM representations of the original signal.

The TEO approach was first proposed by Teager [32] and further investigated by Kaiser [15]. Significant research on the theory and applications of the TEO operator has been conducted during the past 15 years.

Its long-term properties have been studied in detail in [20], [21], [26] and for noisy signals in [2] and [3]. Its AM-FM demodulation capabilities have been compared in [26] with those of the classic linear integral approach of the Hilbert transform or of TEO-inspired instantaneous FM tracking schemes based on adaptive linear prediction [11], [31]. The applications of TEO include speech analysis [6], [21], [27], robust feature extraction for speech recognition [7], [8], communications [30], and image texture analysis [16], [18]. So far, the majority of the analysis in this area has mainly dealt with the properties of TEO-based demodulation algorithms and not with the operator itself. Additionally, the short-and medium-term properties of the TEO have not been formally investigated. In this paper, we investigate the properties of the TEO as a function of the window length. Furthermore, we compare the TEO's performance with that of the SEO for the problem of short-term energy estimation in additive noise. However, the effects of bandpass filtering[1] on the short-time energy estimation process is not addressed here, for more information see [9].

The main contributions of this paper include:

(i) The TEO and SEO performance is investigated for short and medium-length analysis windows. It is shown that performance is a function of the window length. It also depends on the signal and noise spectral characteristics.

(ii) The approximation of the noise with a discrete harmonic model is proposed, significantly simplifying the noisy signal energy analysis and offering insight into the operators' behavior.

(iii) The relationship between signal differentiation and energy estimation is presented. Under certain conditions, the energy operators' performance is improved when they are applied to the signal's time-derivatives.

(iv) The effect of discrete-time sampling on the the performance of the energy operator is investigated. This effect becomes significant when the signal has high frequency content and the sampling frequency is comparable to the Nyquist rate.

The proposed analysis provides some general guidelines on selecting the appropriate energy operator with respect to the minimization of the short-term energy estimation error. This error is primarily based on the spectral characteristics of the signal and noise, as well as, on the analysis window length.

This paper is organized as follows: In Section II, the clean AM-FM and the harmonic noise models are introduced. In this context, the long-term average properties of the TEO and SEO are presented.

[1]The TEO gives meaningful results only if applied to narrowband signals [20]. Henceforth, both clean and noise signals are considered either as narrowband or as pre-processed via narrowband filtering.

Then, the short- and medium-term average energy estimates and their performance are studied in Section III. In Section V a similar analysis is performed for discrete-time signals. The application of the energy operators to the signal derivatives is investigated in Section IV. The effects of discrete-time sampling on the energy estimation scheme are examined in Section VI. Finally, experimental results for short-term energy computation of synthetic and real speech signals are presented in Sections VII and VIII. The overall conclusions are provided in Section IX.

## II. Performance of Energy Operators in Noise

### A. Signal and Noise Model

Consider the narrowband input noisy signal

$$y(t) = x(t) + v(t) \tag{3}$$

where $x(t)$ are the desired clean and $v(t)$ the uncorrelated noise signal, respectively. Herein, we use a narrowband amplitude-frequency modulation (AM-FM) model for the clean signal:

$$x(t) = a(t)\cos(\phi_x(t)) \tag{4}$$

where $\phi_x(t) = \int_0^t \omega_x(\tau)d\tau + \theta_x$,

$$\omega_x(t) = \frac{d\phi_x(t)}{dt}$$

and $a(t)$ are the instantaneous frequency and amplitude signals, and $\theta_x$ is a phase offset. The underlying assumption of the AM-FM model is that both information signals $a(t),\ \omega_x(t)$ do not vary too fast or too greatly compared to the carrier frequency.

The noise signal $v(t)$ is approximated by a sum of $K$ stationary sinusoids $v_i(t)$ with fixed amplitudes $b_i$, frequencies $\omega_i$ and random phase offsets $\theta_i$:

$$v(t) = \sum_{i=1}^{K} b_i \cos\left(\phi_i(t)\right), \quad \phi_i(t) = \omega_i t + \theta_i \tag{5}$$

where each random phase offset $\theta_i$ is uniformly distributed over $[-\pi,\ \pi]$, and the component frequencies are assumed distinct, i.e., $\omega_i \neq \omega_j$ for $i \neq j$. An assumption for independent, identically distributed (i.i.d.) phase offsets is only necessary for the results presented in Section III and Appendix II; i.e., all the major theoretical results hold true for arbitrary phase values. In general, the proposed model (5) can approximate a wide range of known noise models when the amplitude and phase parameters are appropriately chosen [24].

### *B. TEO-Based Noisy Energy Estimation*

By applying the TEO to the noisy signal $y(t)$ and ignoring, henceforth, the time index $t$ for notational simplicity, we obtain (see also [3])

$$\Psi[y] = \Psi[x] + \Psi[v] + \underbrace{2\dot{x}\dot{v} - \ddot{x}v - x\ddot{v}}_{Cross-Terms} \tag{6}$$

Thus, the TEO output of the noisy signal is the sum of the individual signal and noise Teager energies plus some cross-terms. Applying $\Psi$ to the AM-FM signal yields

$$\Psi[x] = \left(\dot{a}\cos(\phi_x) - a\omega_x \sin(\phi_x)\right)^2 - a\cos(\phi_x) \cdot$$

$$\cdot \left(\ddot{a}\cos(\phi_x) - 2\dot{a}\omega_x \sin(\phi_x) - a\dot{\omega}_x \sin(\phi_x) - a\omega_x^2 \cos(\phi_x)\right)$$

Assuming that $a(t)$ varies slowly so that $\Psi[a] \approx 0$, (as shown in [20])

$$\Psi[x] \approx (a\omega_x)^2 + \frac{1}{2}a^2\dot{\omega}_x \cos(2\phi_x) \tag{7}$$

According to [3], [20], the long-term time-average $<\Psi[x]>$ is given by[2]

$$<\Psi[x]> \approx <a^2\omega_x^2> \tag{8}$$

where the quantity $<f(t)>$ for an arbitrary signal $f(t)$ is defined as the signal time-average

$$<f(t)> \triangleq \frac{1}{T}\int_0^T f(t)dt \tag{9}$$

and $T$ is the duration of the analysis window. In the case of window lengths $T$ smaller than the smallest signal period (with respect to its spectral content), this equation provides the short-term average. When $T$ exceeds the largest signal period (or equivalently $T \to +\infty$), the $<\cdot>$ shall imply long-term averages. Henceforth, if it is not otherwise stated, we shall assume that the long-term averages are estimated.

By applying the TEO to the noise (5), we obtain

$$\Psi[v] = \sum_i (b_i\omega_i)^2 + \frac{1}{2}\sum_i \sum_{j\neq i} b_i b_j \omega_i(\omega_i + \omega_j)\cos(\phi_i - \phi_j)$$

$$+ \frac{1}{2}\sum_i \sum_{j\neq i} b_i b_j \omega_i(\omega_i - \omega_j)\cos(\phi_i + \phi_j) \tag{10}$$

[2]In [20], the instantaneous frequency signal is modeled as $\omega_x(t) = \omega_c + q(t)$, where $\omega_c$ is its center frequency and $q(t)$ a zero-mean signal fluctuating around the center frequency. By considering all assumptions about $q(t)$ presented in [20], it follows that the long-term time-average $< \cos(2\phi_x(t)) >$ is approximately zero.

where $i, j = 0, .., K-1$. Its time-average is

$$<\Psi[v]> \approx \sum_i b_i^2 \omega_i^2 \tag{11}$$

The rest of the cross-terms (of $\Psi[v]$) consist of sums of cosines with different amplitude and frequency values, thus their long-term time-averages equal to zero [3]. Denoting the cross-terms of $\Psi[y(t)]$, (6), as

$$\Psi_{cross}[x, v] = 2\dot{x}\dot{v} - x\ddot{v} - \ddot{x}v \tag{12}$$

and substituting the signal representations of (4) and (5) yields

$$\begin{aligned}
&\Psi_{cross}[x, v] = \\
&\frac{1}{2}\sum_i [b_i \left(a\omega_x^2 + a\omega_i^2 - \ddot{a} - 2a\omega_x\omega_i\right) \cdot \cos(\phi_x + \phi_i) + \\
&b_i \left(a\omega_x^2 + a\omega_i^2 - \ddot{a} + 2a\omega_x\omega_i\right) \cdot \cos(\phi_x - \phi_i) + \\
&b_i \left(2\dot{a}\omega_x + a\dot{\omega}_x - 2\dot{a}\omega_i\right) \cdot \sin(\phi_x + \phi_i) + \\
&b_i \left(2\dot{a}\omega_x + a\dot{\omega}_x + 2\dot{a}\omega_i\right) \cdot \sin(\phi_x - \phi_i)]
\end{aligned}$$

For a slowly varying $a(t)$, the $\Psi_{cross}[x, v]$ is approximated by

$$\begin{aligned}
\Psi_{cross}[x, v] \approx \\
&\frac{1}{2}\sum_i ab_i[(\omega_x - \omega_i)^2 \cos(\phi_x + \phi_i) + (\omega_x + \omega_i)^2 \cos(\phi_x - \phi_i) \\
&+ \dot{\omega}_x \left(\sin(\phi_x + \phi_i) + \sin(\phi_x - \phi_i)\right)]
\end{aligned} \tag{13}$$

By similar reasoning as above, $<\Psi_{cross}[x, v]> \approx 0$ (it is shown analytically in Appendix I for the case of a sinusoid signal $x(t)$). Thus, the average Teager energy of the noisy signal is given by

$$<\Psi[y]> \approx <a^2\omega_x^2> + \sum_i b_i^2\omega_i^2 \tag{14}$$

The *normalized TEO deviation* $\mathcal{D}_\mathcal{T}$ is defined as the ratio of the difference between the noisy and clean energy estimates over the clean estimate:

$$\mathcal{D}_\mathcal{T}(y) = \frac{<\Psi[y]> - <\Psi[x]>}{<\Psi[x]>} \approx \frac{\sum_i b_i^2\omega_i^2}{<a^2\omega_x^2>} \tag{15}$$

The difference $<\Psi[y]> - <\Psi[x]>$ always takes non-negative values for long-term analysis of narrowband signals. However, no such guarantees exist for wideband signals, where the approximation in (14) is not applicable. In such cases, one might choose, instead, to compute the absolute value of the normalized TEO deviation.

### C. SEO-Based Noisy Energy Estimation

Applying the SEO to the noisy signal

$$S[y] = x^2 + v^2 + 2xv = S[x] + S[v] + S_{cross}[x, v] \tag{16}$$

where $S_{cross}[x, v]$ are the SEO cross-terms. Substituting the clean and noise signals,

$$S[x] = S_d[x] + S_e[x] = \frac{1}{2}a^2 + \frac{1}{2}a^2\cos(2\phi_x) \tag{17}$$

$$S[v] = \frac{1}{2}\sum_i b_i^2(1 + \cos(2\phi_i)) \tag{18}$$

$$\begin{aligned} S_{cross}[x, v] &= \sum_i ab_i\left(\cos(\phi_x + \phi_i) + \cos(\phi_x - \phi_i)\right) \\ &+ \frac{1}{2}\sum_i\sum_{j\neq i} b_i b_j\left(\cos(\phi_i + \phi_j) + \cos(\phi_i - \phi_j)\right) \end{aligned} \tag{19}$$

where $S_d[x] = \frac{1}{2}a^2$ and $S_e[x] = \frac{1}{2}a^2\cos(2\phi_x)$ are the desired and error components of $S[x]$, respectively.

For the reasons stated in the analysis of $\Psi_{cross}[x, v]$, it holds that $<S_{cross}[x, v]>\approx 0$, $<\cos(2\phi_x)>\approx 0$, $<\cos(2\phi_i)>\approx 0$. Thus, the long-term averaged SEO estimate $<S[y]>$ is given by

$$<S[y]>\approx \frac{1}{2}<a^2> + \frac{1}{2}\sum_i b_i^2 \tag{20}$$

and the *normalized SEO deviation* $\mathcal{D}_\mathcal{S}$ is given by[3]

$$\mathcal{D}_\mathcal{S}(y) = \frac{<S[y]> - <S_d[x]>}{<S_d[x]>} \approx \frac{\sum_i b_i^2}{<a^2>} \tag{21}$$

Henceforth, the signal index will be ignored in $\mathcal{D}_\mathcal{T}$ and $\mathcal{D}_\mathcal{S}$, for notational simplicity.

Using Parseval's theorem[4] [22], the normalized SEO deviation $\mathcal{D}_\mathcal{S}$ can be expressed as:

$$\mathcal{D}_\mathcal{S} = \frac{\sum_i b_i^2}{\int_B |X(\omega)|^2 d\omega}$$

where $X(\omega)$ is the Fourier Transform of the clean signal and the integral is evaluated within the frequency band of interest $B$. Similarly, using relations presented in [5], [29], the normalized TEO deviation $\mathcal{D}_\mathcal{T}$ can be expressed in the frequency domain as:

$$\mathcal{D}_\mathcal{T} = \frac{\sum_i b_i^2\omega_i^2}{\int_B \omega^2|X(\omega)|^2 d\omega}$$

[3]Note that $<S[x]>$ can be used instead of $<S_d[x]>$ in (21) because $<S_e[x]>\approx 0$ for long-term averaging. For (very) short-time averages, however, the term $<S_e[x]>$ becomes relevant as detailed in Section III-B.

[4]The equations dictated by the Parseval Theorem are theoretically valid only when infinite time has elapsed, otherwise a finite-length window should be introduced. Herein, we assume that the window length is long enough to enable the omission of such windows from the equations.

The TEO deviation can be seen as the ratio of the second order spectral centroid of noise over the signal [23], [29], while, the SEO deviation is the ratio of the zero-th order spectral centroids. The SEO and TEO deviations are approximately equal, i.e., $\mathcal{D}_{\mathcal{S}} \approx \mathcal{D}_{\mathcal{T}}$, when: (i) the signal and noise occupy the same very narrow frequency band, or (ii) the signal and noise have very similar spectral profiles (ideally scaled version of each other). In general, when the noise is concentrated in frequencies lower than those of the signal, the TEO outperforms the SEO and vice-versa. Examples elucidating these phenomena and the performance of the energy operators are presented in Section VII.

## III. Medium-Term and Short-Time Properties of Energy Operators

The analysis presented in the previous section assumes that the duration of the averaging window is long enough to ignore all transient deviation terms. Next, the performance of the energy operators is analyzed for different window lengths, namely: (i) Medium-term analysis: The highpass transient terms can be ignored but not the lowpass terms that have not been fully averaged out and thus, contribute to the estimation error, and (ii) Short-term analysis: All transient terms (both highpass and lowpass) contribute to the estimation error and should be taken into account in the analysis. The terms "medium-term" and "short-term" do not correspond to a fixed range of window duration $T$. The actual short-term and medium-term range is determined by the spectral content of the signal (and noise). For example, for a 100 Hz sinusoid, the short-term range would be approximately from 0 to 10 ms (one period of the signal), and the mid-range from 10 to 100 ms.

In general, the normalized TEO and SEO deviations can be separated into three components: (i) the *long term deviation*, as in (6) and (19), (ii) the *lowpass deviation* component that consists of sinusoidal terms corresponding to differences of frequencies, henceforth referred to as ${\mathcal{D}_{\mathcal{T}}}^{-}$ and ${\mathcal{D}_{\mathcal{S}}}^{-}$, respectively, and (iii) the *highpass deviation* component consisting of sinusoids with angular frequencies equal to the sums of the individual component frequencies, henceforth referred to as ${\mathcal{D}_{\mathcal{T}}}^{+}$ and ${\mathcal{D}_{\mathcal{S}}}^{+}$

$$\mathcal{D}_{\mathcal{T}} = \frac{\sum_i {b_i}^2 {\omega_i}^2}{<a^2 \omega_x^2>} + {\mathcal{D}_{\mathcal{T}}}^{-} + {\mathcal{D}_{\mathcal{T}}}^{+} \tag{22}$$

$$\mathcal{D}_{\mathcal{S}} = \frac{\sum_i {b_i}^2}{<a^2>} + {\mathcal{D}_{\mathcal{S}}}^{-} + {\mathcal{D}_{\mathcal{S}}}^{+} \tag{23}$$

Next, we analyze the behavior of the lowpass and highpass transient terms assuming that $x(t)$ is a sinusoid, i.e., $a(t) = a =$ constant, and $\omega_x(t) = \omega_x =$ constant. The following analysis is based on the results derived in Appendices I and II.

### A. Medium-Term Time Average Properties

The lowpass transient terms are given by

$$\mathcal{D}_{\mathcal{T}}{}^{-} = \sum_i \frac{(\omega_x+\omega_i)^2}{2\omega_x^2}\frac{b_i D_{xi}}{aT(\omega_x-\omega_i)} + \sum_i\sum_{j\neq i}\frac{\omega_i(\omega_i+\omega_j)}{2\omega_x^2}\frac{b_i b_j D_{ij}}{a^2T(\omega_i-\omega_j)} \tag{24}$$

$$\mathcal{D}_{\mathcal{S}}{}^{-} = \sum_i \frac{2b_i D_{xi}}{aT(\omega_x-\omega_i)} + \sum_i\sum_{j\neq i}\frac{b_i b_j D_{ij}}{a^2T(\omega_i-\omega_j)} \tag{25}$$

where $D_{ij}$ contains sinusoids with frequencies $\omega_i-\omega_j$, as defined in Appendix I. A direct correspondence exists between the two terms in $\mathcal{D}_{\mathcal{T}}{}^{-}$ and $\mathcal{D}_{\mathcal{S}}{}^{-}$. Based on the assumption that $\omega_i$, $\omega_j$ are in the vicinity of $\omega_x$, then $(\omega_x+\omega_i)^2/(2\omega_x^2)\approx 2$ and $\omega_i(\omega_i+\omega_j)/(2\omega_x^2)\approx 1$. Thus, the first order approximation gives:

$$\mathcal{D}_{\mathcal{T}}{}^{-} \approx \mathcal{D}_{\mathcal{S}}{}^{-} \tag{26}$$

and the TEO and SEO performance is similar for medium-length windows. When the spectral content of the noise is symmetrically distributed around $\omega_x$ then[5] $\mathcal{D}_{\mathcal{T}}{}^{-} = \mathcal{D}_{\mathcal{S}}{}^{-}$. However, when the spectral content of the noise is mostly concentrated over frequencies lower than $\omega_x$, the medium-term performance of the TEO is better than that of the SEO (and vice versa for noise at frequencies higher than $\omega_x$). Thus, the relative medium-term TEO and SEO performance appears quite similar to the corresponding long-term performance of these operators.

### B. Short-Time Average Properties

The highpass transient terms equal to

$$\mathcal{D}_{\mathcal{T}}{}^{+} = \sum_i \frac{(\omega_x-\omega_i)^2}{2\omega_x^2}\frac{b_i S_{xi}}{aT(\omega_x+\omega_i)} + \sum_i\sum_{j\neq i}\frac{\omega_i(\omega_i-\omega_j)}{2\omega_x^2}\frac{b_i b_j S_{ij}}{a^2T(\omega_i+\omega_j)} \tag{27}$$

$$\mathcal{D}_{\mathcal{S}}{}^{+} = \sum_i \frac{2b_i S_{xi}}{aT(\omega_x+\omega_i)} + \sum_i\sum_{j\neq i}\frac{b_i b_j S_{ij}}{a^2T(\omega_i+\omega_j)} + \frac{S_{xx}}{2T\omega_x} + \sum_i \frac{b_i^2 S_{ii}}{2a^2T\omega_i} \tag{28}$$

[5]A fine detail to be noted here is that for $\omega_i=\omega_x+d$ the TEO deviation is larger, while the opposite is true for $\omega_i=\omega_x-d$. When the sum of these deviations is computed, the TEO deviation will be slightly higher than that of the SEO because the TEO deviation relation is quadratic with frequency. The result is most noticeable for large bandwidths, both for medium- and long-term.

where $S_{ij}$ contains sinusoids with frequencies $\omega_i + \omega_j$, as defined in Appendix I. There is a direct correspondence between the first two terms of ${\mathcal{D}_\mathcal{T}}^+$ and ${\mathcal{D}_\mathcal{S}}^+$; however, ${\mathcal{D}_\mathcal{S}}^+$ contains two additional terms. Given that $\omega_i$, $\omega_j$ are in the vicinity of $\omega_x$, as above, it follows that $(\omega_x - \omega_i)^2/(2\omega_x^2) \ll 1$ and $\omega_i|\omega_i - \omega_j|/(2\omega_x^2) \ll 1$. Thus, the values of ${\mathcal{D}_\mathcal{T}}^+$ are much smaller than those of ${\mathcal{D}_\mathcal{S}}^+$, on average. Formally, for small values of $T$, it holds that

$$E\{({\mathcal{D}_\mathcal{S}}^+)^2\} \gg E\{({\mathcal{D}_\mathcal{T}}^+)^2\} \tag{29}$$

where $E(.)$ denotes expectation over the random phases of signal and noise. The mean square normalized deviation values are analytically estimated in Appendix II, assuming that the noise component phases are i.i.d. uniformly distributed. For all the reasons stated above, the short-term TEO performance is expected to be better than that of the SEO. It is, also, important to note that all terms in ${\mathcal{D}_\mathcal{T}}^+$ and ${\mathcal{D}_\mathcal{S}}^+$ are inversely proportional to the frequency content, i.e., the frequency $\omega_x$. Consequently, for smaller frequency values, the deviation terms are further emphasized.

In the general case of AM-FM signals, conclusions similar to the above can be derived, since the deviation terms share the same form. However, the time-varying nature of the signals increases the complexity of the analysis and the mathematical simplicity of the results cannot be reached.

## IV. Applying Energy Operators to Signal Derivatives

In this section, the performance of the energy operators applied to signal derivatives is evaluated, and interesting analogies are drawn between the long-term behavior of the TEO and SEO. The $\ell^{th}$-order time derivative $x^{(\ell)}(t)$ of the AM-FM signal $x(t)$ defined in (4) can be approximated by [3]

$$d^\ell x(t)/dt^\ell \triangleq x^{(\ell)}(t) \approx a(t)\omega_x^\ell(t)\cos\left(\phi_x(t) + \ell\frac{\pi}{2}\right) \tag{30}$$

By applying the TEO on $x^{(\ell)}(t)$, we get

$$\Psi\left[x^{(\ell)}\right] \approx a^2\omega_x^{2(\ell+1)} \tag{31}$$

as shown in Appendix III. Following the same steps outlined in (6)-(15) for the $0^{th}$ derivative case, the averaged TEO output of the $\ell^{th}$-order time derivative of the noisy signal is

$$<\Psi\left[y^{(\ell)}\right]> \approx <a^2\omega_x^{2(\ell+1)}> + \sum_i b_i^2\omega_i^{2(\ell+1)} \tag{32}$$

and the normalized TEO deviation defined as in (15) can be approximated by

$${\mathcal{D}_\mathcal{T}}(y^{(\ell)}) \approx \frac{\sum_i b_i^2\omega_i^{2(\ell+1)}}{<a^2\omega_x^{2(\ell+1)}>} \tag{33}$$

Similarly, the long-term average SEO energy of $y^{(\ell)}(t)$ is

$$<S[y^{(\ell)}]>\approx \frac{1}{2}\left(<a^2\omega_x^{2\ell}> + \sum_i b_i^2\omega_i^{2\ell}\right) \tag{34}$$

and the normalized SEO deviation

$$\mathcal{D}_{\mathcal{S}}(y^{(\ell)}) \approx \frac{\sum_i b_i^2\omega_i^{2\ell}}{<a^2\omega_x^{2\ell}>} \tag{35}$$

Comparing the long-term performance of the TEO and SEO in terms of normalized deviation, shown in (33) and (35), respectively, it is clear that the TEO applied to the $(\ell-1)^{th}$ signal derivative $y^{(\ell-1)}(t)$ performs equivalently to the SEO applied to the $\ell^{th}$ signal derivative $y^{(\ell)}(t)$. This is experimentally verified in Section VII-B. However, for very short-term averaging, the performance of the TEO remains superior to that of the SEO as discussed in Section III-B.

To better understand the behavior of the TEO (or SEO) applied to high-order time derivatives of a noisy signal, note the $\omega^{2(\ell+1)}$ frequency weighting term in the numerator and denominator of (33). The normalized TEO deviation according to (33) is equal to the ratio of the $2(\ell+1)$-order noise spectral centroid over that of the signal. Thus, for noise that is spectrally concentrated at frequencies well below those of the signal, the normalized TEO deviation decreases[6] with $\ell$. Overall, the short-, medium- and long-term qualitative behavior of TEO (and SEO) outlined in Sections II and III holds also for the signal derivatives, although, the effects are amplified by additional frequency weighting.

## V. Performance of Discrete-Time Energy Operators in Noise

The discrete-time signals are derived by sampling the corresponding continuous-time ones for $t = nT_s$,

$$\begin{aligned} x[n] &= A[n]\cos(\Phi_x[n]) \\ v[n] &\approx \sum_{i=1}^{K} v_i[n] = \sum_{i=1}^{K} B_i\cos(\Phi_i[n]) \end{aligned} \tag{36}$$

where $T_s$ is the sampling period and $A[n] = a(nT_s)$, $B_i = b_i =$ constant, $\Phi_x[n] = \phi_x(nT_s)$, $\Phi_i[n] = \phi_i(nT_s)$. As proposed in [20], [21] for the time-differentiation operation $d\Phi_x[n]/dn$, the integer time index $n$ is symbolically treated as a continuous variable. That is,

$$\Omega_x[n] \triangleq \omega_x(nT_s)\cdot T_s \quad \text{and} \quad \Omega_i[n] \triangleq \omega_i \cdot T_s \tag{37}$$

Finally, the noise-corrupted discrete-time signal is represented by $y[n] = x[n] + v[n]$.

[6]Although the TEO deviation decreases with $\ell$, the desired term $<a^2\omega_x^{2(\ell+1)}>$ also becomes increasingly frequency weighted, a potentially undesired effect.

Complementary to the continuous-time domain analysis of Sections II and III, a noisy energy analysis for the corresponding discrete-time signals is presented next. The *discrete-time squared energy operator (DSEO)* is defined, following (1), as $S\left[x[n]\right] \triangleq x^2[n]$. Further, the *discrete-time Teager-Kaiser energy operator (DTEO)* is given, when the TEO time-derivatives are approximated by one-sample differences [21], by

$$\Psi^d[x[n]] \triangleq \left(x^2[n] - x[n+1] \cdot x[n-1]\right)/T_s^2 \tag{38}$$

Applying the DTEO to the noisy discrete signal gives

$$\Psi^d[y[n]] = \Psi^d[x[n]] + \Psi^d[v[n]] + \Psi^d_{cross}[x[n], v[n]] \tag{39}$$

where the DTEO cross-terms are

$$\begin{aligned}\Psi^d_{cross}[x[n], v[n]] = (2x[n]v[n] - x[n+1]v[n-1] \\ -x[n-1]v[n+1])/T_s^2 = \\ = \sum_i (2A[n]B_i \cos(\Phi_x[n]) \cdot \cos(\Phi_i[n]) - \\ A[n+1]B_i \cos(\Phi_x[n+1]) \cdot \cos(\Phi_i[n-1]) - \\ A[n-1]B_i \cos(\Phi_x[n-1]) \cdot \cos(\Phi_i[n+1]))/T_s^2\end{aligned} \tag{40}$$

The $\Psi^d_{cross}[x[n], v[n]]$ terms consist of products of cosines with phases $\Phi_x,\ \Phi_i$. Therefore, their long-term averages approximately equal zero, similarly to the results obtained for the continuous-time case in Section II. So

$$<\Psi^d[y]> \approx <\Psi^d[x]> + <\Psi^d[v]> \tag{41}$$

where $<\Psi^d[x]>,\ <\Psi^d[v]>$ are the averaged clean and noise discrete-time TEO energies, respectively. The first term is approximated [20], [21] by

$$<\Psi^d[x[n]]> = \frac{<A^2[n]\sin^2(\Omega_x[n])>}{T_s^2} \approx \frac{<A^2[n]\Omega_x^2[n]>}{T_s^2} \tag{42}$$

The average noise DTEO output is approximated by

$$<\Psi^d[v[n]]> \approx \frac{1}{T_s^2}\sum_i B_i^2\Omega_i^2 \tag{43}$$

By combining (41)-(43), we obtain[7]

$$<\Psi^d[y[n]]> \approx \frac{1}{T_s^2}\left(<A^2[n]\Omega_x^2[n]> + \sum_i B_i^2\Omega_i^2\right) \tag{44}$$

[7]The approximation is exact when $T_s \to 0$. In general, the approximation error is small under certain conditions detailed in Section VI.

Thus, the *discrete-time DTEO deviation* ${\mathcal{D}_\mathcal{T}}^d$ is given by

$$ {\mathcal{D}_\mathcal{T}}^d(y[n]) = \frac{\sum_i B_i^2 \Omega_i^2}{<A^2[n]\Omega_x^2[n]>} \tag{45} $$

similarly to the continuous-time case.

The discrete-time analysis concerning the squared energy operator (DSEO) is straightforward,

$$ S[y[n]] = S[x[n]] + S[v[n]] + S_{cross}[x[n], v[n]] $$

where

$$ S[x[n]] = \frac{1}{2} A^2[n] \left(1 + \cos(2\Phi_x[n])\right) \tag{46} $$

$$ S[v[n]] = \frac{1}{2} \sum_i B_i^2 \left(1 + \cos(2\Phi_i[n])\right) \tag{47} $$

and

$$ S_{cross}\left[x[n], v[n]\right] = 2 \sum_i A[n] B_i \cos(\Phi_x[n]) \cos(\Phi_i[n]) + \sum_i \sum_{j \neq i} B_i B_j \cos(\Phi_i[n]) \cos(\Phi_j[n]) \tag{48} $$

The long-term averages of all DSEO cross-term can be approximated by $<S_{cross}\left[x[n], v[n]\right]> \approx 0$, as stated above. Thus, the long-term averaged DSEO output is given by

$$ <S\left[y[n]\right]> \approx \frac{1}{2} \left( <A^2[n]> + \sum_i B_i^2 \right) \tag{49} $$

and the *discrete-time DSEO deviation* ${\mathcal{D}_\mathcal{S}}^d$ is

$$ {\mathcal{D}_\mathcal{S}}^d(y[n]) = \frac{\sum_i B_i^2}{< A^2[n] >} \tag{50} $$

${\mathcal{D}_\mathcal{T}}^d$ and ${\mathcal{D}_\mathcal{S}}^d$ can be considered as the discrete-time approximations of the continuous-time deviations, (45) and (50) (this holds true for the case of short- and medium-length analysis windows too, however, these results are not further elaborated here due to lack of space). The sampling process greatly affects the DTEO energy estimation process via the approximations made. In this context, the underlying phenomena hereby described are independent of the sampling period $T_s$ only under certain conditions, detailed in Section VI. Finally, equations similar to those in Section IV can be obtained for the DTEO and DSEO when applied to high-order derivatives of the discrete-time signal (approximated as differences).

## VI. Discrete Time TEO Approximation Error

The discretization of the TEO introduces an *approximation error* due to the use of one-sample differences. The DTEO approximation error $\Delta$ evaluated at $t = nT_s$ is

$$\Delta \triangleq \left(a(t)\omega_x(t)\right)^2|_{t=nT_s} - A^2[n]\sin^2\left(\Omega_x[n]\right)/T_s^2 \Rightarrow$$

$$\Delta = a^2(nT_s)\left[\omega_x^2(nT_s) - \sin^2\left(\omega_x(nT_s)\cdot T_s\right)/T_s^2\right]$$

The quality of the approximation depends on the product $\omega_x(nT_s)\cdot T_s$. In the limiting case, where $\omega_x(nT_s)$ tends to $0$ the approximation error also tends to $0$, because

$$\lim_{\omega_x \to 0} \sin\left(\omega_x(nT_s)\cdot T_s\right) = \omega_x(nT_s)\cdot T_s$$

Assuming that $\omega_x(t) = \omega_c + q(t)$, where $\omega_c$ is the center frequency and $q(t)$ a slow-varying signal, the product $\omega_c T_s$ determines the quality of the approximation. Thus, when processing a signal though a filterbank, the approximation will be better for low frequency bands than for the high frequency ones. In addition, the approximation error can be reduced by increasing the sampling frequency.

The quality of the discrete-time approximation is also affected by the input signal's derivative order. Consider the Taylor series expansion for a sinusoid

$$\sin(\omega(t)) \approx \omega(t) - \frac{\omega^3(t)}{6} \tag{51}$$

where the first term is the desired one and the second term is a rough estimate of the approximation error. The discretization of the TEO is based on the assumption that

$$\omega_x^2(nT_s) \approx \sin^2\left(\Omega_x[n]\right)/T_s^2$$

Similarly, when the TEO is applied on time-derivatives of the signal the discrete-time approximation is[8]

$$\omega_x^{2(\ell+1)}(nT_s) \approx \sin^{2(\ell+1)}\left(\Omega_x[n]\right)/T_s^{2(\ell+1)} \tag{52}$$

Thus, the *normalized approximation error* $D^{(\ell)}_{DTEOapprox.}$ of the DTEO applied to the $\ell^{th}$ derivative of the signal is

$$D^{(\ell)}_{DTEOapprox.} \approx \frac{\left(\omega_x - \frac{\omega_x^3}{6}\right)^{2(\ell+1)} - \omega_x^{2(\ell+1)}}{\omega_x^{2(\ell+1)}} \tag{53}$$

[8]By considering the DTEO definition and its one-sample differences one may write

$$\Psi^d\left[\frac{d^\ell x[n]}{dm^\ell}\right] \approx \Psi^d\left[x[n] - x[n-\ell]\right]$$

This approximation is used here instead of the one proposed in (52); both approximations yield similar results [2], [3].

The normalized approximation error for higher-order derivatives can be also expressed as follows:

$$D^{(\ell)}_{DTEOapprox.} \approx (\ell+1) \cdot D^{(0)}_{DTEOapprox.} \tag{54}$$

i.e., the normalized approximation error increases linearly with the derivative order. Overall, for low sampling frequencies, high signal carrier frequencies and/or high-order signal derivatives the approximation error of the DTEO becomes large, as experimentally verified in Section VII. Note that better discrete-time approximations have been proposed in the literature [4], [12] and can be used to overcome some of the DTEO approximation errors.

## VII. Experiments with Synthetic Signals

Next, the proposed energy estimation methods are applied to simple synthetic signals, namely, pure sinusoids in additive white noise. For pure sinusoids the energy deviation is directly computable and the validity of the theoretical results can be experimentally verified.

Consider three sinusoids with center frequencies $100,\ 150$ and $200$ Hz and phase offset $\pi/4$, corrupted by additive (bandpassed) white noise. The sinusoids were sampled at 2 kHz, resulting in the discrete signals $x_1[n] = \cos\left(\frac{10}{100}\pi n + \frac{\pi}{4}\right)$, $x_2[n] = \cos\left(\frac{15}{100}\pi n + \frac{\pi}{4}\right)$, $x_3[n] = \cos\left(\frac{20}{100}\pi n + \frac{\pi}{4}\right)$. The white noise signal was bandpass filtered by a Finite Impulse Response (FIR) filter with 201 coefficients and passband in the interval $[100,\ 200]$ Hz. A total of 1000 instances of the bandpassed white noise signal $v[n]$ were randomly generated and added to the pure sinusoids to create 1000 instances of the noisy signals $y_j[n] = x_j[n] + v[n],\ j = 1, 2, 3$, with Signal to Noise Ratio (SNR) 0 dB.

The noise signal $v[n]$ can be modeled by $K = 100$ sinusoid signals $v_i[n]$ $(i = 1, \cdots, 100)$ with frequencies linearly distributed over the passband and random phases $\theta_i$ uniformly distributed over the interval $[-\pi,\ \pi]$, as in (36). The noise amplitude coefficients $B_i$ should be equal and normalized to ensure $\mathrm{SNR} = 0$ dB. The noise signal can then be approximated by

$$v[n] \approx \sum_{i=1}^{K} \frac{1}{\sqrt{K}} \cos\left(\left(\frac{\pi}{10} + \frac{i}{K}\frac{\pi}{10}\right) n + \theta_i\right) \tag{55}$$

### A. Short-Time Energy of Noisy Sinusoidal Signals

The theoretical long-term values of the normalized deviations ${\mathcal{D}_{\mathcal{T}}}^d(y_j)$ and ${\mathcal{D}_{\mathcal{S}}}^d(y_j)$ were computed using (45) and (50). The theoretically computed ${\mathcal{D}_{\mathcal{T}}}^d$ values were

$$ {\mathcal{D}_{\mathcal{T}}}^d(y_1) = \frac{1}{K} \frac{\sum_{i=1}^{K} \left( \frac{\pi}{10} + \frac{i}{K} \frac{\pi}{10} \right)^2}{(\pi/10)^2} \approx 2.34 $$

$$ {\mathcal{D}_{\mathcal{T}}}^d(y_2) \approx 1.04 $$

$$ {\mathcal{D}_{\mathcal{T}}}^d(y_3) \approx 0.58 $$

Similarly, the DSEO normalized deviation is

$$ {\mathcal{D}_{\mathcal{S}}}^d(y_j) = 1, \;\; j = 1, 2, 3 $$

The DTEO and DSEO short-term energy was experimentally estimated using 1000 instances of $y_j[n]$. The *root mean square*[9] (rms) and *standard deviation* values (std) of the DTEO and DSEO normalized deviation were experimentally computed and compared with their theoretical values. The results are presented for a 500 ms averaging window in Table I. Good agreement (typically within one standard deviation of the rms value) is achieved between the theoretical and experimental results. Small differences observed between the theoretical and experimental values can be attributed to: (i) the approximation of time-derivatives with one-sample differences, and (ii) the approximation of narrowband white noise in (55). It is interesting to note that the DSEO outperforms the DTEO in terms of normalized deviation for $y_1[n]$, and vice versa for $y_3[n]$.

The experimentally computed RMS deviations ${\mathcal{D}_{\mathcal{T}}}^d, \; {\mathcal{D}_{\mathcal{S}}}^d$ are shown in Fig. 1(a)-(c) as functions of the analysis window duration $T$ that takes values between 0 and 500 ms. In Fig. 1(d)-(f), the results are shown when the experiment was repeated with the phases of the sinusoids $x_j[n]$ taking random values (uniformly) in the interval $[-\pi, \; \pi]$. Again RMS deviations are shown, averaged over 1000 noisy signal instances as a function of $T$. In all plots, transient phenomena fade out as the window length $T$ increases and the normalized deviations ${\mathcal{D}_{\mathcal{T}}}^d$ and ${\mathcal{D}_{\mathcal{S}}}^d$ converge to their long-term values. A detailed analysis of the transient error terms is presented in Appendices I and II.

Table I and Fig. 1 verify the basic conclusions drawn by the theoretical analysis. Specifically, the DSEO significantly outperforms the DTEO for noisy signal $y_1[n]$, as shown in Fig. 1(a),(d). This is expected because the clean signal energy is concentrated at 100 Hz, while the noise energy content is placed at higher frequencies (spread between 100 and 200 Hz with an average approx. at 150 Hz). The opposite

[9]The experimentally computed rms value can be compared with the mean square deviation analytically derived in Appendix II.

holds true for the case of $y_3[n]$, where the signal energy is now placed at a higher frequency, i.e., 200 Hz, (see Fig. 1(c),(f)). Finally, for $y_2[n]$ where the clean and noise signals present similar average spectral characteristics the medium- and long-term average performance of the DTEO and DSEO is comparable, as shown in Fig. 1(b) and (e).

For very-short term analysis ($T$<5 ms), the DTEO performance is always superior to that of DSEO, regardless of the signals' spectral content, due to the transient effects outlined in Section III-B. Also, the medium-term behavior (up to 100 ms approximately) of the DTEO and DSEO is similar to their long-term behavior, as predicted in Section III-A. Finally, the DTEO and DSEO performance is not affected much by the phase of the signal and noise, as can be seen by a direct comparison of Fig. 1(a),(d), 1(b),(e) and 1(c),(f).

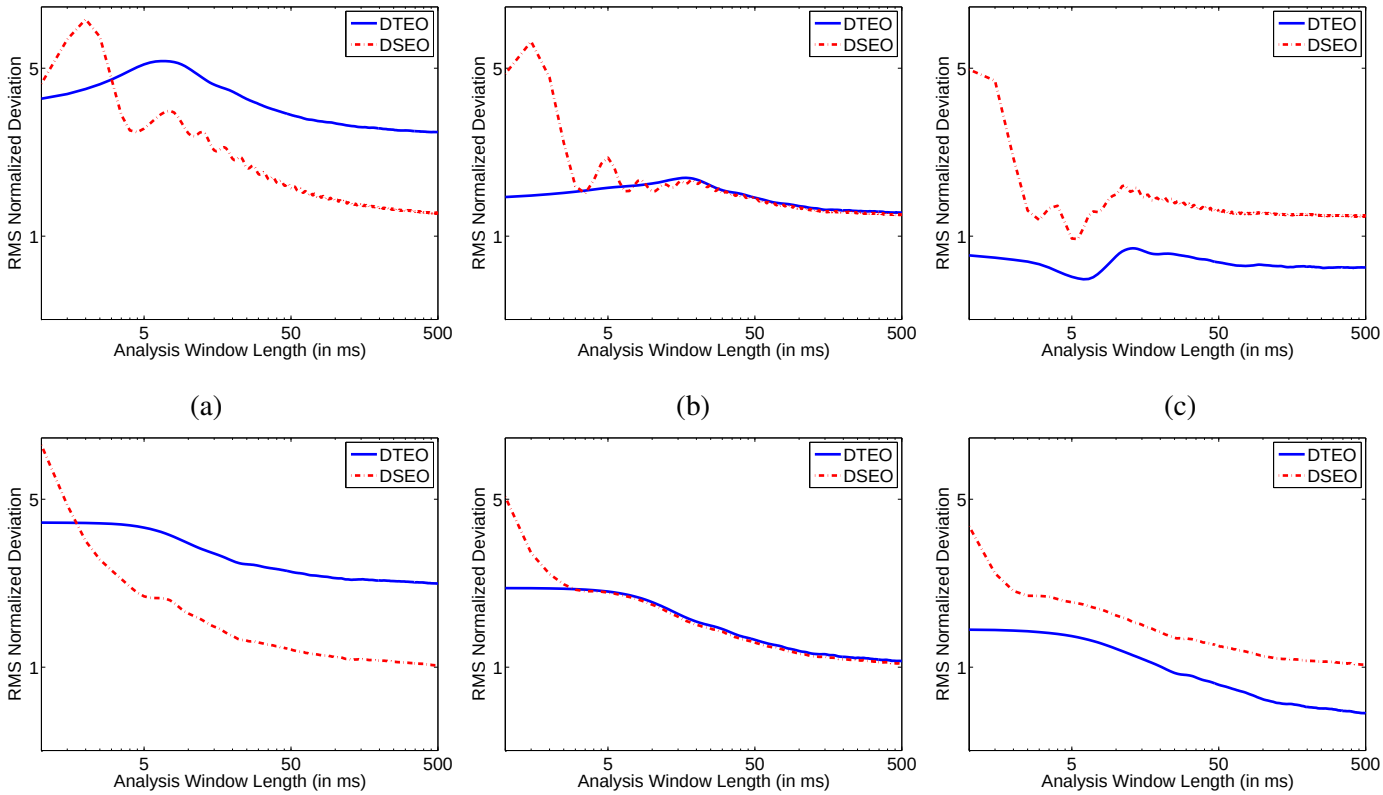


Fig. 1. DTEO and DSEO RMS normalized deviations ${\mathcal{D}_{\mathcal{T}}}^d$, ${\mathcal{D}_{\mathcal{S}}}^d$, as a function of window length $T$ (in ms) for the signals: (a) $y_1[n]$, (b) $y_2[n]$ and (c) $y_3[n]$. Same for random phase sinusoids in (d)-(f). Deviations shown in all plots are averaged over 1000 instances of the random signals $y_j[n]$. The SNR level is 0 dB. Both x- and y-axis are in log-scale.

| DTEO and DSEO Normalized Deviation | | | | | | |
|---|---|---|---|---|---|---|
| | $y_1 = x_1 + v$ | | $y_2 = x_2 + v$ | | $y_3 = x_3 + v$ | |
| | rms | std | rms | std | rms | std |
| $\Psi$-Operator | 2.46 | 0.27 | 1.14 | 0.13 | 0.68 | 0.08 |
| Theoretical Value | 2.34 | | 1.04 | | 0.58 | |
| S-Operator | 1.11 | 0.12 | 1.10 | 0.11 | 1.11 | 0.12 |
| Theoretical Value | 1.00 | | 1.00 | | 1.00 | |

TABLE I

DTEO AND DSEO RMS NORMALIZED DEVIATIONS (AND STANDARD DEVIATION OF ESTIMATE) COMPUTED OVER 1000 INSTANCES OF THE RANDOM SIGNALS $y_1$, $y_2$ AND $y_3$. THE SNR LEVEL IS 0 dB AND THE ANALYSIS WINDOW LENGTH IS 500 ms.

### *B. Short-Time Energy of Signal Derivatives*

Herein, we investigate the DTEO and DSEO performance when higher-order derivatives $y_j^{(\ell)}[n]$ of the input signals are employed, where $j = 1, 2, 3$ are the indices of the noisy sinusoids, as defined in the previous section, and $\ell = 1, 2, 3$ are the first, second and third-order derivatives of those signals. Our goal is to verify the theoretical results in (32) and (34), and to compare with the experimentally computed DTEO and DSEO deviations. In the following experiments, first-order derivatives are approximated by one-sample differences. Higher-order derivatives of order $\ell$ are iteratively estimated using one-sample differences of the $(\ell - 1)$-order derivative.

The experimental setup and result presentation is identical to that of Section VII-A, but here signal derivatives are used. The DTEO and DSEO normalized deviations are computed first theoretically using (32), (34), and then experimentally by averaging over 1000 instances of the noisy input signals. The root mean square (rms) and standard deviation (std) of these deviations (along with the theoretical values) are shown in Table II for a $T = 500$ ms window length. Overall, there is a good agreement between the theoretical and experimental results.

The RMS normalized deviations of the DSEO and the DTEO applied to the signal derivatives $y_j^{(\ell)}[n]$ are shown in Fig. 2, as a function of the averaging window length $T$. Again, all results are in agreement with the theory. The performance of the DSEO applied to the $\ell^{th}$ signal derivative and that of the DTEO applied

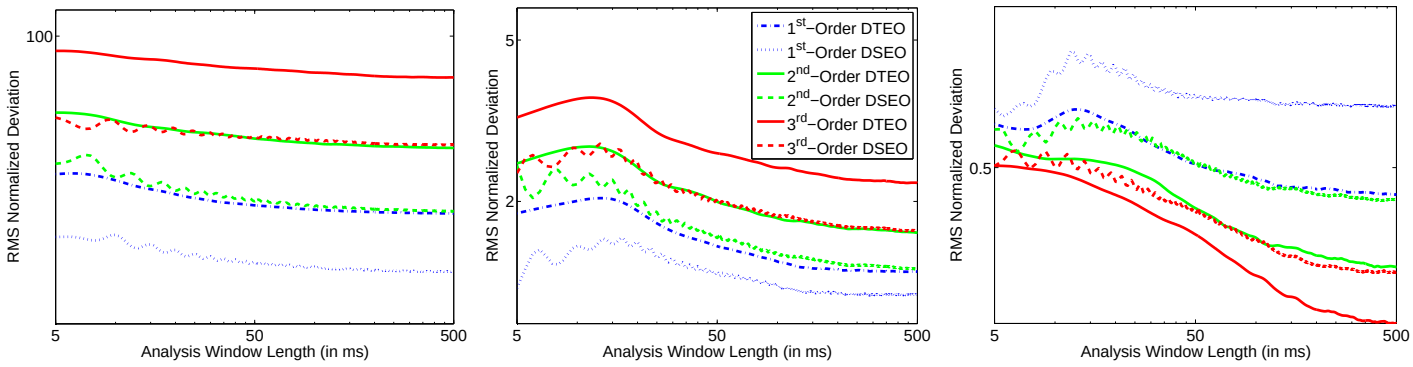


Fig. 2. DTEO and DSEO RMS normalized deviations $\mathcal{D}_{\mathcal{T}}{}^{d}$, $\mathcal{D}_{\mathcal{S}}{}^{d}$, as a function of window length $T$ (in ms) for the signals: (a) $y_1^{(\ell)}[n]$, (b) $y_2^{(\ell)}[n]$ and (c) $y_3^{(\ell)}[n]$, for $\ell = 1, 2, 3$. Deviations shown in all plots are averaged over 1000 instances of the random signals $y_j[n]$. The SNR level is 0 dB. Both x- and y-axis are in log-scale (y-axis range is different in (a)-(c) to enhance readability).

to the $(\ell-1)^{th}$ derivative are very similar for both medium-term[10] and, especially, long-term, as predicted by theory (see also Table II). For the case of $y_3^{(\ell)}[n]$ shown in Fig. 2(c), lower normalized deviations are achieved when high-order derivatives are used, because the signal energy content is concentrated at higher frequencies than the corresponding noise content. The opposite is true for signal $y_1^{(\ell)}[n]$ shown in Fig. 2(a). In general, the normalized deviation of DTEO and DSEO applied to signal derivatives is governed by the amount of frequency weighting as theoretically predicted by (32) and (34).

## VIII. Experiments with Speech Signals

Next, the relative performance of the DTEO and DSEO is evaluated for a realistic speech processing application. The time-frequency distribution of speech signals, in the presence of different types of additive noise, is estimated and the corresponding energy deviations are computed. The proposed filterbank analysis and short-term energy estimation is typically performed by the front-end of a speech recognition system. Our goal is to verify, via these experiments, the theoretical results and to provide further insight in the relative performance of DTEO and DSEO for speech processing applications.

The RMS DTEO and DSEO deviations, defined in (45) and (50), can be interpreted as the *inverse Signal to Noise Ratio* (SNR) where the estimation error is considered as the "noise" and the desired

[10]The very short-term performance of the DTEO and DSEO is not shown in the figure to avoid clutter. As expected, the DTEO significantly outperforms the DSEO for $T < 5$ ms.

| DTEO and DSEO Normalized Deviation | | | | | | |
|---|---|---|---|---|---|---|
| | $(x_1+v)^{(\ell)}$ | | $(x_2+v)^{(\ell)}$ | | $(x_3+v)^{(\ell)}$ | |
| | rms | std | rms | std | rms | std |
| $\ell=1$ | | | | | | |
| $\Psi$-Operator | 6.28 | 0.68 | 1.31 | 0.14 | 0.44 | 0.05 |
| Theoretical Value | 6.22 | | 1.23 | | 0.38 | |
| S-Operator | 2.56 | 0.28 | 1.15 | 0.13 | 0.66 | 0.07 |
| Theoretical Value | 2.34 | | 1.04 | | 0.58 | |
| $\ell=2$ | | | | | | |
| $\Psi$-Operator | 17.41 | 1.92 | 1.63 | 0.18 | 0.32 | 0.04 |
| Theoretical Value | 18.29 | | 1.61 | | 0.29 | |
| S-Operator | 6.53 | 0.72 | 1.32 | 0.15 | 0.43 | 0.05 |
| Theoretical Value | 6.22 | | 1.23 | | 0.38 | |
| $\ell=3$ | | | | | | |
| $\Psi$-Operator | 52.14 | 5.75 | 2.19 | 0.25 | 0.24 | 0.03 |
| Theoretical Value | 57.50 | | 2.24 | | 0.22 | |
| S-Operator | 18.42 | 2.03 | 1.67 | 0.19 | 0.31 | 0.04 |
| Theoretical Value | 18.29 | | 1.61 | | 0.29 | |

TABLE II

DTEO AND DSEO RMS NORMALIZED DEVIATIONS (AND STANDARD DEVIATION OF ESTIMATE) COMPUTED OVER 1000 INSTANCES OF THE FIRST, SECOND AND THIRD ORDER DERIVATIVES OF THE RANDOM SIGNALS $y_1$, $y_2$ AND $y_3$. THE SNR LEVEL IS $0$ dB AND THE ANALYSIS WINDOW LENGTH IS $500$ ms.

energy term as the "signal". Specifically, we define $\text{SNR}_{\mathcal{S}} \triangleq -10\log(\mathcal{D}_{\mathcal{S}}{}^d)$ as the SNR in dBs for the DSEO and similarly $\text{SNR}_{\mathcal{T}} \triangleq -10\log(\mathcal{D}_{\mathcal{T}}{}^d)$ for DTEO. Herein, all results are presented in terms of the *log distortion difference* between the DSEO and DTEO, i.e., $\text{SNR}_{\mathcal{S}} - \text{SNR}_{\mathcal{T}}$ in dBs. Negative distortion difference values indicate better DTEO performance, and vice versa for DSEO.

The DTEO and DSEO values are estimated over speech signals corrupted by various types of additive noise. For this purpose, the NOISEX-92 noise database is used, containing ten typical noise samples, each with different spectral characteristics [33]. These noise signals are down-sampled to 16 kHz and

added to the speech samples[11] extracted from the TIMIT database, while keeping the global average Signal-to-Noise Ratio (SNR) fixed at $\mathrm{SNR} = 5$ dB[12]. The clean speech is used as the reference signal for computing the normalized deviation and the log distortion difference.

In this experiment, only five, i.e., *babble, buccaneer 1, volvo, factory 1 and white* noise types are examined. Specifically: (i) babble noise is acquired when 100 people are recorded speaking in a canteen where individual voices are slightly audible [33], (ii) buccaneer noise is mainly a low frequency type of noise with the addition of a high frequency component, (iii) volvo noise presents mainly a lowpass structure and can be considered stationary, (iv) factory noise was recorded near plate-cutting and electrical welding equipment [33] and it is non-stationary (e.g., contains hammer blows), (v) white noise exhibits equal energy per frequency bin. These noise signals are added to 1000 different instances of the phonemes */aa/, /ae/, /sh/* and */f/*, all extracted from the TIMIT database.

To simulate the filterbanks commonly-used in speech processing applications, a linearly-spaced, Gabor filterbank with 25 filters and fixed 3 dB-bandwidth overlap percentage of $50\%$ is used [6], [8], [28]. Short-term DTEO and DSEO energy estimates are computed for each frequency bin using analysis frames with duration of 30 ms (updated every 10 ms).

The median[13] log distortion difference between the DTEO and DSEO time-frequency estimates is presented in Table III for two voiced (/aa/, /ae/) and two unvoiced phonemes (/sh/, /f/). The median is computed over 1000 instances of each phone, both in time (over all frames) and frequency (over all frequency bins). Overall, the DTEO significantly outperforms the DSEO for all noise types with the exception of white noise. The performance gap is larger for lowpass volvo noise and for the phonemes /sh/, /f/. In general, the DTEO outperforms the DSEO when the *spectral tilt*[14] of the noise is smaller compared to that of the signal, e.g., for lowpass volvo noise or for fricative sounds (where the signal's spectral tilt is rising up to approx. 3 kHz). This observation is consistent with (45), (50), i.e., DTEO is superior when

[11]The noise signals have a duration of approximately 235 sec, so a portion of the noise signal is randomly selected and added to each speech signal.

[12]The SNR value is estimated as the mean ratio of the speech over the noise signal energies per frame. Then, the noise signals are scaled so that the global mean SNR is $5$ dB. Therefore, this value refers to the wide-band speech signal and suggests that the SNR level is, on average, 5 dBs.

[13]We use the median instead of the root mean square estimate here to get rid of outliers. For certain time-frequency bins, the energy of the signal is too low resulting in very large normalized deviation values.

[14]The *spectral tilt* is defined as the slope of a line that best fits the log power spectrum of the input signal, more details can be found in [10].

| Median Log Distortion Difference Between DSEO and DTEO (in dB) for Noisy Speech Phonemes | | | | | |
|---|---|---|---|---|---|
| | Noise Type | | | | |
| Phoneme | Babble | Buccaneer 1 | Volvo | Factory 1 | White |
| /aa/ | -0.06 | -0.03 | -0.44 | -0.06 | 0.05 |
| /ae/ | -0.04 | -0.02 | -0.43 | -0.06 | 0.05 |
| /sh/ | -0.17 | -0.15 | -0.82 | -0.18 | -0.05 |
| /f/ | -0.13 | -0.10 | -0.81 | -0.14 | 0.001 |

TABLE III

MEDIAN LOG DISTORTION DIFFERENCE BETWEEN THE DSEO AND DTEO ESTIMATES COMPUTED OVER ALL SPEECH FRAMES AND FREQUENCY BANDS FOR 1000 INSTANCES (PER PHONEME). RESULTS ARE SHOWN FOR FIVE TYPES OF NOISE AND FOUR TYPES OF PHONEMES. SNR IS 5 dB.

the noise energy is concentrated in lower frequencies than those of the signal. Approximation errors and transient effects also affect performance, as discussed next.

In Fig. 3, the median log distortion difference is shown as a function of the filter index (or equivalently the signal's carrier frequencies) for phonemes /aa/ and /sh/, and for (a) babble and (b) white noise. Two additional conclusions about the relative performance of DTEO and DSEO can be drawn from Fig. 3, namely: (i) The DSEO performs significantly worse than the DTEO for the first few filters. This is due to additional transient error terms of DSEO. As discussed in Section III, the magnitude of the transient terms is inversely proportional to frequency and, thus, the transient terms take large values for the first few filters. (ii) The discrete-time approximation error of DTEO becomes large at high frequencies, as discussed in Section VI. This explains the worse performance of DTEO for the last few filters. Overall, the experimental results are in agreement with the theory and provide important intuition about the DTEO and DSEO performance for speech processing applications.

## IX. CONCLUSIONS

In this paper, the properties of the Teager-Kaiser and the squared energy operators in the presence of additive noise are examined as a function of the short-term averaging window length. This analysis covers both the continuous- and discrete-time domains. Furthermore, the robustness of the energy estimation

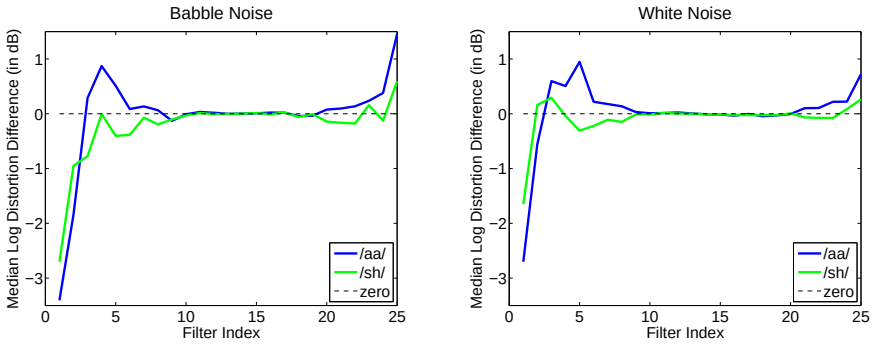


Fig. 3. Median of the log distortion differences between the DSEO and DTEO as a function of filter index for different noise types: (a) babble, and (b) white. The global signal SNR is equal to 5 dB. The median is computed over 1000 instances of the phonemes /aa/ and /sh/. The filterbank consists of 25 Gabor filters, linearly spaced with fixed overlap. Negative values indicate better DTEO performance.

process is investigated when the TEO and SEO are applied to the derivatives (or differences) of the original signal. Overall, we have concluded that the following factors affect the TEO and SEO performance as short-term energy estimators: (i) *The relative differences between the spectral shape of the signal and noise*, or more specifically, the ratio of the second spectral centroid of the noise over that of the signal. In general, the TEO outperforms the SEO when the noise is more "lowpass" than the signal, and vice versa. (ii) *The duration of the analysis window*: the TEO outperforms the SEO for short analysis windows ($<5$ ms). For all other cases, the clean and noise spectra must be considered. (iii) *The magnitude of the short- and medium-term transient error terms is inversely proportional to the signals' frequency content*: transient phenomena are more prominent for signals with low frequency components, especially for the SEO that contains two additional transient terms. (iv) *The sampling frequency*: the discrete-time approximation error of the DTEO increases when the center (average) signal and noise frequencies move towards the Nyquist frequency. In addition, we have shown that more robust energy estimates may be obtained by applying the operators to the high-order derivatives of the signal[15] for noise with "lowpass" spectral characteristics (compared to those of the signal). In this context, the long-term properties of the SEO applied to the $\ell^{th}$ signal derivative are equivalent to those of the TEO applied to the $(\ell-1)^{th}$ signal derivative (baring DTEO approximation errors).

The results are experimentally verified on synthetic and real speech signals. Based on preliminary results

[15]The estimated energy is weighted by the frequency, an unwanted side-effect. Also, approximation errors creep up in discrete-time implementations.

using such signals we can state that, in general, the TEO appears to be more robust than the SEO for speech-related applications. The results in this paper can be exploited for a variety of signal processing applications where short-term energy estimation in noise is required, such as, telecommunication and image processing applications. In general, for applications where the noise spectral characteristics are known (and differ from those of the signal), a short-time energy estimator exhibiting optimal performance can be selected based on the results of this paper.

## APPENDIX I
## SHORT-TERM TEAGER-KAISER AND SQUARED ENERGY ESTIMATION FOR SINUSOIDS IN ADDITIVE NOISE

In this section, the short-term average energy of a sinusoid $x(t) = a \cdot \cos(\omega_x t + \theta_x)$ corrupted by additive noise $v(t)$ is computed. The energy of the noisy signal $y(t) = x(t) + v(t)$ is estimated using the squared energy and Teager-Kaiser operators over a time window of duration $T$. The short-time average of the TEO $<\Psi[y]>$ is

$$<\Psi[y]>= \frac{1}{T}\left[\int_0^T \Psi[x]dt + \int_0^T \Psi[v]dt + \int_0^T \Psi_{cross}[x,v]dt\right]$$

Given that $<\Psi[x]>= (a\omega_x)^2$, and based on (10)

$$<\Psi[v]>= \frac{1}{T}\int_0^T \Psi[v]dt = \frac{1}{T}\sum_i (b_i\omega_i)^2 \int_0^T dt + \frac{1}{2T}\int_0^T \sum_i \sum_{i\neq j} b_i b_j \omega_i(\omega_i+\omega_j)\cos(\phi_i-\phi_j)dt + \frac{1}{2T}\int_0^T \sum_i \sum_{i\neq j} b_i b_j \omega_i(\omega_i-\omega_j)\cos(\phi_i+\phi_j)dt \quad (56)$$

Let us define

$$S_{ij} = \sin\left[(\omega_i+\omega_j)T + (\theta_i+\theta_j)\right] - \sin\left(\theta_i+\theta_j\right) \quad (57)$$

$$D_{ij} = \sin\left[(\omega_i-\omega_j)T + (\theta_i-\theta_j)\right] - \sin\left(\theta_i-\theta_j\right) \quad (58)$$

then the short-time average of the noise is

$$<\Psi[v]>= \frac{1}{T}\int_0^T \Psi[v]dt = \sum_i (b_i\omega_i)^2 + \sum_i \sum_{i\neq j} \frac{b_i b_j}{2T}\omega_i\left(\frac{\omega_i-\omega_j}{\omega_i+\omega_j}S_{ij} + \frac{\omega_i+\omega_j}{\omega_i-\omega_j}D_{ij}\right) \quad (59)$$

Similarly, the short-time average of the TEO cross-terms is

$$<\Psi_{cross}[x,v]>=\frac{1}{T}\int_0^T \Psi_{cross}[x,v]dt = \sum_i \frac{ab_i}{2T}\left(\frac{(\omega_x-\omega_i)^2}{\omega_x+\omega_i}S_{xi}+\frac{(\omega_x+\omega_i)^2}{\omega_x-\omega_i}D_{xi}\right) \quad (60)$$

where $S_{xi},\ D_{xi}$ are defined as in (57), (58). The normalized deviation $\mathcal{D}_{\mathcal{T}}$ defined in (15), is given by

$$\mathcal{D}_{\mathcal{T}}(y)=\frac{<\Psi[v]>+<\Psi_{cross}[x,v]>}{<\Psi[x]>}$$

Similarly for the SEO,

$$<S[y]>=<S_d[x]>+<S_e[x]>+<S[v]>+<S_{cross}[x,v]>$$

From (17)-(19),

$$<S_d[x]>=\frac{1}{T}\int_0^T \frac{1}{2}a^2dt=\frac{a^2}{2} \quad (61)$$

$$<S_e[x]>=\frac{1}{T}\int_0^T \frac{1}{2}a^2\cos(2\phi_x)dt=\frac{a^2}{4T\omega_x}S_{xx} \quad (62)$$

$$<S[v]>=\frac{1}{2T}\sum_i b_i^2\int_o^T (1+\cos(2\phi_i))\,dt = = \sum_i \frac{b_i^2}{2}+\sum_i \frac{b_i^2}{4T\omega_i}S_{ii} \quad (63)$$

$$<S_{cross}[x,v]>=\sum_i \frac{ab_i}{T}\left(\frac{S_{xi}}{\omega_x+\omega_i}+\frac{D_{xi}}{\omega_x-\omega_i}\right)+ \sum_i\sum_{j\neq i}\frac{b_ib_j}{2T}\left(\frac{S_{ij}}{\omega_i+\omega_j}+\frac{D_{ij}}{\omega_i-\omega_j}\right) \quad (64)$$

where $S_{xx}$, $S_{ii}$, $S_{xi}$ are defined as in (57), and $D_{xi}$ is defined as in (58).

From (21), the normalized deviation $\mathcal{D}_{\mathcal{S}}$ is given by

$$\mathcal{D}_{\mathcal{S}}(y)=\frac{<S_e[x]>+<S[v]>+<S_{cross}[x,v]>}{<S_d[x]>}$$

The deviations $\mathcal{D}_{\mathcal{T}}$ and $\mathcal{D}_{\mathcal{S}}$ contain both lowpass and highpass terms, e.g., $D_{ij}$ and $S_{ij}$, correspondingly. There is a direct correspondence between the TEO and SEO error terms, however, the SEO has two additional highpass error terms containing the quantities $S_{xx}$ and $S_{ii}$. In addition, both the desired and error terms of TEO are multiplied by additional frequency squared terms (compared to the SEO), e.g., $\omega_x^2$, $(\omega_x\pm\omega_i)^2$. The additional highpass terms in SEO result is significantly higher error compared to the TEO for very short-term energy estimation.

All TEO and SEO error terms contain the $1/T$ multiplicative term, i.e., the magnitude of both lowpass and highpass transient phenomena is inversely proportional to the analysis window length $T$. Thus, as the analysis window length $T$ increases, the RMS normalized deviations $\mathcal{D}_{\mathcal{T}}$ and $\mathcal{D}_{\mathcal{S}}$ converge to their long-term averaging values, namely, $\frac{\sum_i (b_i\omega_i)^2}{(a\omega_x)^2}$, and $\frac{\sum_i b_i^2}{a^2}$, respectively.

## APPENDIX II
## MEAN SQUARE ENERGY ESTIMATION ERROR FOR RANDOM PHASE SINUSOIDS IN ADDITIVE NOISE

In this section, both $x(t) = a \cdot \cos(\omega_x t + \theta_x)$ and $v(t) = \sum_i b_i \cos(\omega_i t + \theta_i)$ are assumed random signals with $\theta_x$, $\theta_i$ being independent random variables uniformly distributed over the interval $[-\pi, \pi]$. Next, the expected values of the squared normalized TEO and SEO deviations, i.e., $E\{{\mathcal{D}_{\mathcal{T}}}^2\}$ and $E\{{\mathcal{D}_{\mathcal{S}}}^2\}$ respectively, are computed.

Given i.i.d random variables $\theta_i$, $\theta_j$ uniformly distributed in $[-\pi, \pi]$, the random variables $\theta_i + \theta_j$, $\theta_i - \theta_j$ are also i.i.d. and follow the symmetric triangular distribution in $[-2\pi,\ 2\pi]$. It follows that the random variables $S_{ij},\ D_{ij}$ defined in (57), (58) exhibit the properties

$$E\{S_{ij}\} = 0 \quad \text{and} \quad E(D_{ij}) = 0 \tag{65}$$

$$E\{S_{ij}S_{kl}\} = \begin{cases} 1 - \cos[(\omega_i + \omega_j)T], \text{ if } i = k, j = l \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{66}$$

$$E\{D_{ij}D_{kl}\} = \begin{cases} 1 - \cos[(\omega_i - \omega_j)T], \text{ if } i = k, j = l \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{67}$$

$$E\{S_{ij}D_{kl}\} = 0 \tag{68}$$

for any i.i.d. random variables $\theta_i$, $\theta_j$, $\theta_k$, $\theta_l$, uniformly distributed in $[-\pi,\ \pi]$.

Based on (65)-(68), the mean square normalized deviation of the TEO is computed[16],

$$E\{{\mathcal{D}_{\mathcal{T}}}^2(y)\} = \frac{E\{<\Psi[v]>^2 + <\Psi_{cross}[x,v]>^2\}}{<\Psi[x]>^2}$$

because the expected value of the mean square error product term $<\Psi[v]><\Psi_{cross}[x,v]>$ is zero, and the denominator does not depend on the (random) phase. The expected value of the first term is

$$E\{<\Psi[v]>^2\} = \left(\sum_i (b_i\omega_i)^2\right)^2 + \sum_i \sum_{j\neq i} \frac{b_i^2 b_j^2 \omega_i^2}{4T^2} \cdot$$

$$\cdot \left[\left(\frac{\omega_{ij}^-}{\omega_{ij}^+}\right)^2 (1 - \cos(\omega_{ij}^+ T)) + \left(\frac{\omega_{ij}^+}{\omega_{ij}^-}\right)^2 (1 - \cos(\omega_{ij}^- T))\right]$$

[16]The numerator of $E\{{\mathcal{D}_{\mathcal{T}}}^2(y)\}$ is the mean square error.

and, similarly, for the second term

$$E\{<\Psi_{cross}[x,v]>^2\} = \sum_i \frac{a^2 b_i^2}{4T^2}.$$
$$\cdot \left[ \frac{\left(\omega_{xi}^-\right)^4}{\left(\omega_{xi}^+\right)^2}(1-\cos(\omega_{xi}^+ T)) + \frac{\left(\omega_{xi}^+\right)^4}{\left(\omega_{xi}^-\right)^2}(1-\cos(\omega_{xi}^- T)) \right]$$

where we have defined $\omega_{ij}^+ = \omega_i + \omega_j$, $\omega_{ij}^- = \omega_i - \omega_j$ to simplify notation.

The mean square normalized deviation of the SEO is

$$E\{{\mathcal{D}_{\mathcal{S}}}^2(y)\} = \frac{E\{<S_e[x]>^2\} + <S[v]>^2 + <S_{cross}[x,v]>^2\}}{<S_d[x]>^2}$$

because the expected value of all product terms is equal to zero, and the denominator does not depend on the phase. Based on (65)-(68), the three terms in the numerator are equal to

$$E\{<S_e[x]>^2\} = \frac{a^4}{16T^2\omega_x^2}(1-\cos(2\omega_x T))$$
$$E\{<S[v]>^2\} = \left(\sum_i \frac{b_i^2}{2}\right)^2 + \sum_i \frac{b_i^4}{16T^2\omega_i^2}(1-\cos(2\omega_i T))$$

$$E\{<S_{cross}[x,v]>^2\} =$$
$$\sum_i \frac{a^2 b_i^2}{T^2}\left[\frac{1-\cos(\omega_{xi}^+ T)}{(\omega_{xi}^+)^2} + \frac{1-\cos(\omega_{xi}^- T)}{(\omega_{xi}^-)^2}\right] +$$
$$\sum_i \sum_{j\neq i} \frac{b_i^2 b_j^2}{4T^2}\left[\frac{1-\cos(\omega_{ij}^+ T)}{(\omega_{ij}^+)^2} + \frac{1-\cos(\omega_{ij}^- T)}{(\omega_{ij}^-)^2}\right]$$

The expected values of the desired TEO and SEO terms do not depend on the random phases and are given by

$$E\{<\Psi[x]>^2\} = <\Psi[x]>^2 = (a\omega_x)^4$$

and

$$E\{<S_d[x]>^2\} = <S_d[x]>^2 = \frac{a^4}{4}$$

The transient error terms of the SEO and TEO can be grouped in two categories, i.e., those that contain sums of frequencies $(1-\cos(2\omega_i T))$, $(1-\cos(2\omega_x T))$, $(1-\cos(\omega_{ij}^+ T))$ and $(1-\cos(\omega_{xi}^+ T))$, that dominate for very small averaging windows $T$, and those that contain differences of frequencies $(1-\cos(\omega_{ij}^- T))$, $(1-\cos(\omega_{xi}^- T))$ and dominate for medium-size averaging windows. The two additional terms in $E\{{\mathcal{D}_{\mathcal{S}}}^2(y)\}$, namely, $(1-\cos(2\omega_i T))$, $(1-\cos(2\omega_x T))$, are the cause of the poor performance of the SEO for very small averaging windows $T$. Finally, the transient terms of the mean square error decrease as $1/T^2$ for both the TEO and the SEO.

## APPENDIX III
## ESTIMATING DTEO AND DSEO FOR SIGNAL DERIVATIVES

Using the approximation

$$x^{(\ell)} \approx a(t)\left(\omega(t)\right)^{\ell}\cos\left(\phi_x(t)+\ell\frac{\pi}{2}\right)$$

proposed in [3], where $x(t)$ is defined in (4) and $\ell = 0, 1, \ldots$ as in (30), yields

$$\Psi[x^{(\ell)}] = \left(x^{(\ell+1)}\right)^2 - x^{(\ell)}x^{(\ell+2)} \approx$$
$$\approx a^2\omega^{2\ell+2}\sin^2\left(\phi(t)+\ell\frac{\pi}{2}\right) + a^2\omega^{2\ell+2}\cos^2\left(\phi(t)+\ell\frac{\pi}{2}\right)$$

Thus

$$\Psi[x^{(\ell)}] \approx a^2\omega^{2(\ell+1)} \tag{69}$$

Similarly, for the SEO operator we have

$$S[x^{(\ell)}] = a^2\omega^{2\ell}\cos^2\phi \Rightarrow$$
$$S[x^{(\ell)}] = \frac{1}{2}a^2\omega^{2\ell} + \frac{1}{2}a^2\omega^{2\ell}\cos(2\phi) \tag{70}$$

## REFERENCES


[1] R. G. Baraniuk, "Beyond Time-Frequency Analysis: Energy Densities in One and Many Dimensions", *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2305-2314, Sept. 1998.

[2] A. C. Bovik, J. P. Havlicek, M. D. Desai and D. S. Harding, "Limits on Discrete Modulated Signals", *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 867-879, Apr. 1997.

[3] A. C. Bovik, P. Maragos and T. F. Quatieri, "AM-FM Energy Detection and Separation in Noise Using Multiband Energy Operators", *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3245-3265, Dec. 1993.

[4] B. Carlsson, A. Ahlen and M. Sternad, "Optimal Differentiation Based on Stochastic Signal Models", *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 341-353, Feb. 1991.

[5] L. Cohen, "Time-Frequency Distributions - A Review", *Proc. IEEE*, vol. 77, no. 7, pp. 941-981, July 1989.

[6] D. Dimitriadis and P. Maragos, "Continuous Energy Demodulation Methods and Application to Speech Analysis", *Speech Commun.*, vol. 48, no. 7, pp. 819-837, July 2006.

[7] D. Dimitriadis, P. Maragos and A. Potamianos, "Robust AM-FM Features for Speech Recognition", *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 621-624, Sept. 2005.

[8] D. Dimitriadis, P. Maragos and A. Potamianos, "Auditory Teager Energy Cepstrum Coefficients for Robust Speech Recognition", in *Proc. 9th Eur. Conf. Speech Commun. Technol.*, 2005, Lisbon, Portugal.

[9] J. Fang and L. E. Atlas, "Quadratic Detectors for Energy Estimation", *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2582-2594, Nov. 1995.

[10] G. Fant , "The Voice Source in Connected Speech", *Speech Commun.*, vol. 22, no. 2-3, pp. 125-139, Aug. 1997.

[11] L. B. Fertig and J. H. McClellan, "Instantaneous Frequency Estimation Using Linear Prediction With Comparisons to the DESAs", *IEEE Signal Process. Lett.*, vol. 3, pp. 54-56, Feb. 1996.

[12] P. Flajoleta and R. Sedgewick, "Mellin Transforms and Asymptotics: Finite Differences and Rice's Integrals", *Theoretical Computer Science*, vol. 144, no. 1-2, pp. 101-124, June 1995.

[13] S. Gazor and W. Zhang, "Speech Probability Distribution", *IEEE Signal Process. Lett.*, vol. 10, pp. 204-207, July 2003.

[14] J. F. Kaiser, "Some Observations on Vocal Tract Operation from a Fluid Flow Point of View", *Vocal Fold Physiology: Bio-mechanics, Acoustics and Phonatory Control*, I. R. Titze and R. C. Scherer (Eds.), Denver Center for Performing Arts, Denver, CO, pp. 358-386, 1983.

[15] J. F. Kaiser, "On a Simple Algorithm to Calculate the 'Energy' of a Signal", in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, 1990, Albuquerque, NM, pp. 381-384.

[16] I. Kokkinos, G. Evangelopoulos and P. Maragos,"Texture Analysis and Segmentation Using Modulation Features, Generative Models and Weighted Curve Evolution", *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 31, no. 1, pp. 142-157, Jan. 2009.

[17] S. Lu and P. C. Doerschuk, "Nonlinear Modeling and Processing of Speech Based on Sums of AM-FM Formant Models", *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 773-782, Apr. 1996.

[18] P. Maragos and A. C. Bovik, "Image Demodulation Using Multidimensional Energy Separation", *J. Opt. Soc. Amer.*, vol. 12, no. 9, pp. 1867-1876, 1995.

[19] P. Maragos and A. Potamianos, "Higher Order Differential Energy Operators", *IEEE Signal Process. Lett.*, vol. 2, no. 8, pp. 152-154, Aug. 1995.

[20] P. Maragos, J. F. Kaiser and T. F. Quatieri, "On Amplitude and Frequency Demodulation Using Energy Operators", *IEEE Trans. Signal Process.*, vol. 41, no. 4, pp. 1532-1550, Apr. 1993.

[21] P. Maragos, J. F. Kaiser and T. F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis", *IEEE Trans. Signal Process.*, vol. 41, no. 10, pp. 3024-3051, Oct. 1993.

[22] A. V. Oppenheim and R. W. Schafer, "Discrete-Time Signal Processing", 2nd Edition, Prentice Hall: Upper Saddle River, 1999.

[23] K. K. Paliwal, "Spectral Subband Centroid Features for Speech Recognition", in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, 1998, Seattle, WA, pp. 617-620.

[24] A. Papoulis, "Probability, Random Variables and Stochastic Processes", 3rd Edition, McGraw-Hill Inc., 1991.

[25] J. W. Pitton, L. E. Atlas and P. J. Loughlin, "Applications of Positive Time-Frequency Distributions to Speech Processing", *IEEE Trans. Speech and Audio Process.*, vol. 2, no. 4, pp. 554-566, Oct. 1994.

[26] A. Potamianos and P. Maragos, "A Comparison of the Energy Operator and the Hilbert Transform Approach to Signal and Speech Demodulation", *Signal Process.*, vol. 37, no. 1, pp. 95-120, May 1994.

[27] A. Potamianos and P. Maragos, "Speech Formant Frequency and Bandwidth Tracking Using Multiband Energy Demodulation", *J. Acoust. Soc. Amer.*, vol. 99, no. 6, pp. 3795-3806, June 1996.

[28] A. Potamianos and P. Maragos, "Speech Analysis and Synthesis Using an AM-FM Modulation Model", *Speech Commun.*, vol. 28, no. 3, pp. 195-209, July 1999.

[29] A. Potamianos and P. Maragos, "Time-Frequency Distributions for Automatic Speech Recognition", *IEEE Trans. Speech and Audio Process.*, vol. 9, no. 3, pp. 196-200, Mar. 2001.

[30] B. Santhanam and P. Maragos,"Multicomponent AM-FM Demodulation via Periodicity-Based Algebraic Separation and Energy-Based Demodulation", *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 473-490, Mar. 2000.

[31] C. S. Ramalingam, "On the Equivalence of DESA-1a and Prony's Method When the Signal is a Sinusoid", *IEEE Signal Process. Lett.*, vol. 3, no. 5, pp. 141-143, May 1996.

[32] H. M. Teager, "Some Observations on Oral Flow During Phonation", *IEEE Trans. Acoustics, Speech and Signal Process.*, vol. 28, no. 5, pp. 599-601, Oct. 1980.

[33] A. Varga and H. J. M. Steeneken, "Assessment for Automatic Speech Recognition: II. NOISEX-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems", *Speech Commun.*, vol. 12, no. 3, pp. 247-251, July 1993.

**Dimitrios Dimitriadis** (S'99-M'06) received the Diploma degree in ECE and the Ph.D degree both from the National Technical University of Athens, Athens, Greece in 1999 and 2005, respectively.

Since 2005 he has been a postdoctoral Research Associate at the National Technical University of Athens, participating in national and European research projects in the areas of audio and speech processing and recognition. From 2001 to 2002 he was intern at the Multimedia Communications Lab at Bell Labs, Lucent Technologies, Murray Hill, NJ.

His current research interests include speech processing, analysis, synthesis and recognition, multi-modal systems, nonlinear and multi-sensor signal processing.

Dr. Dimitriadis has authored or co-authored over fifteen papers in professional journals and conferences. He is a member of the IEEE Signal Processing Society (SPS) since 1999 and he has served as a reviewer for the IEEE SPS.

**Alexandros Potamianos** (M'92) received the Diploma in ECE from the National Technical University of Athens, Greece in 1990. He received the M.S and Ph.D. degrees in Engineering Sciences from Harvard University, Cambridge, MA, USA in 1991 and 1995, respectively.

From 1991 to June 1993 he was a research assistant at the Harvard Robotics Lab, Harvard University. From 1993 to 1995 he was a research assistant at the Digital Signal Processing Lab at Georgia Tech. From 1995 to 1999 he was a Senior Technical Staff Member at the Speech and Image Processing Lab, AT&T Shannon Labs, Florham Park, NJ. From 1999 to 2002 he was a Technical Staff Member and Technical Supervisor at the Multimedia Communications Lab at Bell Labs, Lucent Technologies, Murray Hill, NJ. From 1999 to 2001 he was an adjunct Assistant Professor at the Department of Electrical Engineering of Columbia University, New York, NY. In the spring of 2003, he joined the Department of Electronics and Computer Engineering at the Technical University of Crete, Chania, Greece as an associate professor.

His current research interests include speech processing, analysis, synthesis and recognition, dialog and multi-modal systems, nonlinear signal processing, natural language understanding, artificial intelligence and multimodal child-computer interaction.

Prof. Potamianos has authored or co-authored over eighty papers in professional journals and conferences. He is the co-author of the paper "Creating conversational interfaces for children" that received a 2005 IEEE Signal Processing Society Best Paper Award; the co-editor of the book "Multimodal Processing and Interaction: Audio, Video, Text". He holds four patents. He is a member of the IEEE Signal Processing Society since 1992 and he is currently serving his second term at the IEEE Speech Technical Committee.

**Petros Maragos** (S'81-M'85-SM'91-F'96) received the EE Diploma from the National Technical University of Athens in 1980, and the M.Sc.E.E. and Ph.D. from Georgia Tech, Atlanta, USA, in 1982 and 1985.

During 1985-1993 he worked as EE professor at the Division of Applied Sciences at Harvard University. In 1993 he joined the ECE faculty at Georgia Tech. During parts of 1996-1998 he was on sabbatical working as director of research at the Institute for Language and Speech Processing in Athens. Since 1998 he has been working as ECE professor at NTUA. His research and teaching interests include signal processing, systems theory, pattern recognition, and their applications to image processing and computer vision, speech and language processing, multimedia, and robotics.

His research has received: a 1987 NSF Presidential Young Investigator Award; a 1988 IEEE SP Society's Young Author Paper Award; a 1994 IEEE SP Senior Award; the 1995 IEEE W.R.G. Baker Prize Award; a 1996 Pattern Recognition Society's Honorable Mention Award; the 2007 EURASIP Technical Achievements Award.

CONTENTS

**I Introduction** 2

**II Performance of Energy Operators in Noise** 4

II-A Signal and Noise Model . . . . . 4

II-B TEO-Based Noisy Energy Estimation . . . . . 5

II-C SEO-Based Noisy Energy Estimation . . . . . 7

**III Medium-Term and Short-Time Properties of Energy Operators** 8

III-A Medium-Term Time Average Properties . . . . . 9

III-B Short-Time Average Properties . . . . . 9

**IV Applying Energy Operators to Signal Derivatives** 10

**V Performance of Discrete-Time Energy Operators in Noise** 11

**VI Discrete Time TEO Approximation Error** 14

**VII Experiments with Synthetic Signals** 15

VII-A Short-Time Energy of Noisy Sinusoidal Signals . . . . . 16

VII-B Short-Time Energy of Signal Derivatives . . . . . 18

**VIII Experiments with Speech Signals** 19

**IX Conclusions** 22

**Appendix I: Short-Term Teager-Kaiser and Squared Energy Estimation for Sinusoids in Additive Noise** 24

**Appendix II: Mean Square Energy Estimation Error for Random Phase Sinusoids in Additive Noise** 26

**Appendix III: Estimating DTEO and DSEO for Signal Derivatives** 28

**References** 28

**Biographies** 30

Dimitrios Dimitriadis . . . . . 30

Alexandros Potamianos . . . . . 31

Petros Maragos . . . . . 31

## LIST OF TABLES

I DTEO and DSEO RMS Normalized Deviations (and Standard Deviation of Estimate) Computed over 1000 Instances of the Random Signals $y_1$, $y_2$ and $y_3$. The SNR level is $0$ dB and the Analysis Window Length is $500$ ms. . . . . 18

II DTEO and DSEO RMS Normalized Deviations (and Standard Deviation of Estimate) Computed over 1000 Instances of the First, Second and Third Order Derivatives of the Random Signals $y_1$, $y_2$ and $y_3$. The SNR level is $0$ dB and the Analysis Window Length is $500$ ms. 20

III Median Log Distortion Difference Between the DSEO and DTEO Estimates computed over all Speech Frames and Frequency Bands for 1000 Instances (per Phoneme). Results are Shown for Five Types of Noise and Four Types of Phonemes. SNR is $5$ dB. . . . . 22

## LIST OF FIGURES

1 DTEO and DSEO RMS normalized deviations ${\mathcal{D}_{\mathcal{T}}}^d$, ${\mathcal{D}_{\mathcal{S}}}^d$, as a function of window length $T$ (in ms) for the signals: (a) $y_1[n]$, (b) $y_2[n]$ and (c) $y_3[n]$. Same for random phase sinusoids in (d)-(f). Deviations shown in all plots are averaged over 1000 instances of the random signals $y_j[n]$. The SNR level is $0$ dB. Both x- and y-axis are in log-scale. . . . . 17

2 DTEO and DSEO RMS normalized deviations ${\mathcal{D}_{\mathcal{T}}}^d$, ${\mathcal{D}_{\mathcal{S}}}^d$, as a function of window length $T$ (in ms) for the signals: (a) $y_1^{(\ell)}[n]$, (b) $y_2^{(\ell)}[n]$ and (c) $y_3^{(\ell)}[n]$, for $\ell = 1, 2, 3$. Deviations shown in all plots are averaged over 1000 instances of the random signals $y_j[n]$. The SNR level is $0$ dB. Both x- and y-axis are in log-scale (y-axis range is different in (a)-(c) to enhance readability). . . . . 19

3 Median of the log distortion differences between the DSEO and DTEO as a function of filter index for different noise types: (a) babble, and (b) white. The global signal SNR is equal to $5$ dB. The median is computed over 1000 instances of the phonemes /aa/ and /sh/. The filterbank consists of 25 Gabor filters, linearly spaced with fixed overlap. Negative values indicate better DTEO performance. . . . . 23