

On the Development of Variable Length Teager Energy Operator (VTEO)

Vikrant Tomar, Hemant A. Patil

Dhirubhai Ambani Institute of Information and Communication Technology, DA-IICT,
Gandhinagar, Gujarat, India

vikrantt@gmail.com, hemant_patil@daiict.ac.in

Abstract

Teager Energy Operator (TEO) proposed by Kaiser and Teager is based on a definition of energy required to generate the signal. TEO gives us the running estimate of energy as a function of amplitude and instantaneous frequency content of the signal. However, it considers three *consecutive* samples to calculate the energy estimate. In this paper, we suggest an alternative and generalized approach to TEO to calculate the instantaneous estimate of the energy where not only *consecutive* but other distant samples can also be incorporated in the calculation of running estimate of the energy and the number of samples taken to calculate energy can also be increased depending on our signal properties to better *capture* the energy content variations in the speech signal.

Index Terms: Variable length Teager Energy Operator (VTEO), Summed-over Variable length Teager Energy Operator (S-VTEO)

1. Introduction

The human speech production process can be modeled by two broad ways. One approach is to model the vocal tract structure using a *source-filter model*. This approach assumes that the underlying source of a speaker's identity is coming from the vocal tract configuration (i.e., size and shape) of the articulators and the manner in which a speaker uses his articulators in sound production. An alternative way to characterize speech production is to model the *airflow pattern* in the vocal tract. The underlying concept here is that while the vocal tract articulators do move to configure the vocal tract shape (making cues for a speaker's identity), it is the resulting airflow properties which serve to excite those models (i.e., speech production model such as source-filter model) which a listener will perceive for a particular speaker's voice [1], [2].

Modeling the time-varying vortex flow (due to *aeroacoustic* sound generation in vocal tracts [7]) is a formidable task and Teager devised a simple algorithm which uses a non-linear energy-tracking operator called as Teager Energy Operator (TEO) (in discrete-time) for signal analysis with the supporting observation that hearing is the process of detecting energy. The concept was further extended to the continuous-domain by Kaiser [2], [3], [4], [8]. Recently, Sinder has applied *aeroacoustic* theory to produce a computationally efficient model of vortex sound generation [9].

Kaiser pointed out that when one speaks about the 'energy' in a signal the usual tendency is to talk about the average of the sum of the squares of the magnitude of that signal as energy required, in a certain sense [3], to generate the signal, i.e., $\frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2$. Another commonly used representation is to take the discrete Fourier transform (DFT) of the signal segment and now the squares of the magnitudes of the frequency samples of the computed transform are

assumed to represent the energy in the respective frequency components, i.e., Parseval's equivalence $\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2$. This is often the case in the speech processing literature where the energy in a speech frame is calculated in this way and hence by this approach of calculating the energy, a unit 10 Hz signal (at a sufficiently large sampling rate) is said to have *same energy* as a unit 1000 Hz signal (of same duration). However the energy required to generate the *acoustic* signal of 10 Hz is much less than that for the 1000 Hz.

Organization of the paper: Section 2 briefly discusses the Teager Energy Operator (TEO) proposed by Kaiser. In section 3 the proposed Variable Length Teager Energy Operator (VTEO) and Summed-over Variable length Teager Energy Operator (S-VTEO) are developed along with their properties. Section 4 discusses experimental results on the proposed VTEO and S-VTEO, and finally section 5 summarizes the paper.

2. Energy measurement and TEO

According to Maragos *et al.*, within one pitch period speech can be modeled as a linear combination of AM-FM signals [5]. Thus Simple Harmonic Motion (SHM) can be used to understand the concept of energy in the context of speech processing. Applying Newton's law of motion to the motion of a mass-spring system having mass m and spring constant k , gives the dynamics as the second order linear differential equation:

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0$$

whose solution is known to be a Simple Harmonic Motion (SHM),

$$x(t) = A \cos(\Omega t + \Phi) \quad (1)$$

where A is amplitude and Ω is angular frequency (in radians) of the oscillation.

The above solution can also be justified in following way. Any periodic function can be decomposed into Fourier series which is an aggregate of an infinite set of sinusoids. The general solution of a second order linear differential equation of this type with positive values of k and m is $\exp(\pm j\sqrt{(k/m)})$ which finally leads to the form $x(t) = A \cos(\Omega t + \Phi)$ [4].

The total energy is the sum of the potential energy in the spring and kinetic energy of the mass and is,

$$E = P.E. + K.E. = \frac{1}{2}kx^2 + \frac{1}{2}m\left(\frac{dx}{dt}\right)^2,$$

substituting $x(t) = A \cos(\Omega t + \Phi)$ and using trigonometry,

$$E = \frac{1}{2}mA^2\Omega^2$$

$$\text{Or, } E \propto A^2 \Omega^2 \quad (2)$$

Eq. (2) is the true energy required to generate $x(n)$ in SHM.

2.1. Development of TEO

Kaiser proposed following algorithm to calculate the running estimate of the energy content in the signal or energy required to generate the signal [3]. In discrete time domain Eq. (1) can be expressed as:

$$x(n) = A \cos(\omega n + \Phi) \quad (3)$$

from (3), we can write:

$$x(n+1) = A \cos(\omega(n+1) + \Phi) \quad (4)$$

$$\text{and, } x(n-1) = A \cos(\omega(n-1) + \Phi) \quad (5)$$

multiplying (4) and (5) and using trigonometry,

$$\begin{aligned} x(n+1)x(n-1) &= A^2 \cos(\omega(n+1) + \Phi) \cos(\omega(n-1) + \Phi) \\ &= [A \cos(\omega n + \Phi)]^2 - A^2 \sin^2 \omega \end{aligned} \quad (6)$$

using Eq. (3) in Eq. (6),

$$A^2 \sin^2 \omega = x^2(n) - x(n+1)x(n-1) \quad (7)$$

for small values of ω , $\sin \omega \approx \omega$, hence (7) can be written as

$$A^2 \omega^2 \approx x^2(n) - x(n+1)x(n-1) = \psi\{x(n)\} \quad (8)$$

where $\psi\{\cdot\}$ is the TEO which gives running estimate of energy of the discrete time signal $x(n)$.

3. Development of VTEO and S-VTEO

In the above section we discussed the traditional TEO in discrete time sense. A close look at the algorithm shows that it involves nonlinear operations on the signal, i.e., square of the signal and multiplication of the signal samples shifted by one in time back and forth, i.e., $x(n-1)$ and $x(n+1)$, respectively. TEO algorithm gives good estimation of energy when signal has sharp changes in frequency or amplitude, i.e., the difference (zero crossing *in some sense*) between the consecutive samples of the signal is observable. But in cases when the difference between the consecutive samples of the signal is minute, i.e., $x(n+1) \cong x(n) \cong x(n-1)$, the TEO will give zero output which means energy required to generate such samples of the signal is zero but this may not be true in any physical system.

In the generalized algorithm we proposed, the number of samples incorporated in energy estimation can be varied up to i past and i future samples, i.e., $x(n-i)$ and $x(n+i)$, instead of only two adjacent samples. And our algorithm also gives the flexibility to select these samples to calculate the running estimate of energy required to generate the signal.

3.1. Derivation of VTEO

Let $x(n)$ be the samples of the signal representing the motion of the oscillatory body. Therefore

$$x(n) = A \cos(\omega n + \Phi) \quad (9)$$

where ω is the digital frequency in radians/sample and is given by $\omega = 2\pi f/f_s$; where f is the analog frequency and f_s

is the sampling frequency. Φ is the arbitrary initial phase in radians. Thus samples of the same signal shifted in time by index i , with respect to present sample, can be expressed as

$$x(n+i) = A \cos(\omega(n+i) + \Phi) \quad (10a)$$

$$x(n-i) = A \cos(\omega(n-i) + \Phi) \quad (10b)$$

[Assumption: for $i > n$, $x(n-i) = 0$]

multiplying Eqs.(10) and using trigonometry, we obtain,

$$\begin{aligned} x(n+i)x(n-i) &= A^2 \cos(\omega(n+i) + \Phi) \cos(\omega(n-i) + \Phi) \\ &= [A \cos(\omega n + \Phi)]^2 - A^2 \sin^2 i\omega \end{aligned} \quad (11)$$

$$\text{or, } A^2 \sin^2 i\omega = x^2(n) - x(n+i)x(n-i) \quad (12)$$

It is important to note that (13) is exact and unique provided that the value of $i\omega$ is restricted to values less than $\pi/2$, the equivalent of one-fourth of the sampling frequency. Also, for small values of $i\omega$, $\sin i\omega \approx i\omega$. Now if we limit the value of ω to $\omega < \pi/4i$, i.e., $f/f_s < 1/8i$ then the relative error in approximation is negligible. Thus on a high sampling rate (12) results in:

$$i^2 A^2 \omega^2 \approx x^2(n) - x(n+i)x(n-i) \quad (13)$$

In Eq. (13) the LHS is the instantaneous estimate of the energy of the signal multiplied by i^2 . We refer to it as the Variable length Teager Energy Operator, abbreviated as VTEO, and represent it by $\xi_i\{\cdot\}$. Hence

$$\xi_i\{x(n)\} = x^2(n) - x(n+i)x(n-i) \quad (14)$$

The proposed differential operator, VTEO, maps the discrete-time samples of a signal onto a running estimate of the signal's energy. It is noteworthy that it also utilizes three samples like the traditional TEO, one present sample and two time shifted samples which can be any i^{th} past and i^{th} future samples unlike the traditional TEO which restrict us to the adjacent two samples. Eq.(14) gives us a good measure of the energy of the oscillating signal when the sampling rate of the signal is greater than δi times the frequency of oscillation of the signal, i.e., at least $2i$ sample points in each quarter cycle of the sinusoidal oscillation. In the next section we discuss another version of the VTEO.

3.1.1. Summed-over Variable length Teager Energy Operator (S-VTEO)

The proposed energy-tracking operator VTEO can be taken to one step further by utilizing more than three samples at once, for example, working on past i samples and future i samples, altogether, and hence better capturing the variations in frequency in the signal and thus the *dependency* in the samples of speech signal.

For this purpose, we sum over Eq. (14) for the range of i , which gives us a new operator which utilizes the $2i+1$ number of samples. We refer to this new operator as Summed-over Variable length Teager Energy Operator and abbreviate as S-VTEO.

$$\begin{aligned} \xi\{x(n), i\} &= \sum_{k=1}^i \xi_k\{x(n)\} \\ &= i \cdot x^2(n) - \sum_{k=1}^i x(n+k)x(n-k) \end{aligned} \quad (15)$$

It is of extreme importance to note here that i used in Eq. (10) to signify the time index of the samples taken into consideration to calculate VTEO and S-VTEO responses now refers to the *length* or *sample length* of VTEO and S-VTEO.

VTEO and S-VTEO can also be represented in terms of the traditional TEO.

From Eq. (13) and (8),

$$\xi_i\{x(n)\} = i^2 \psi\{x(n)\} \quad (16)$$

and from (15) and (16),

$$\xi\{x(n), i\} = \psi\{x(n)\} \sum i^2 \quad (17)$$

Eqs (14) and (15) are generalized equations for VTEO and S-VTEO respectively and using appropriate values of i these can be derived back to the lower length VTEOs/S-VTEOs and eventually to the traditional TEO, for $i = 1$.

$$\xi\{x(n), 1\} = \xi_1\{x(n)\} = \psi\{x(n)\} \quad (18)$$

It should be noted that TEO, VTEO and S-VTEO are energy tracking operators which give the running estimate of energy and not the absolute value of energy over a frame.

3.2. Properties of VTEO and S-VTEO

The TEO for complex valued signals is derived in [3]. The proposed VTEO and S-VTEO, both, follow the following properties in conjunction with the traditional TEO as mentioned in [3], [6].

1. It is independent of the initial phase in the signal.
2. It is symmetric in the sense that changing $x(n) \rightarrow x(-n)$ or $n \rightarrow -n$ i.e., reversing the signal in time does not change the resulting value.
3. The algorithm is robust even if the signal passes through zero, i.e., $x(n) = 0$ as there is no division operation involved.
4. The algorithm is capable of responding very rapidly because it involves only two multiplications and two subtractions *per sample*.

4. Experimental Results

In this section a relative comparison of TEO, and proposed VTEO and S-VTEO is performed on samples of synthetic and real speech signals to show the effectiveness of the proposed operator. In all the experiments below the sample length of VTEO and S-VTEO is kept to be $i = 5$. Also in all the figures below (figure 1 to figure 6) (a) is the original signal, (b) is TEO output, (c) is VTEO output, and (d) is S-VTEO output.

Figure 1 shows the comparison of output of the traditional TEO, and proposed VTEO and S-VTEO when applied on a sinusoid with mix and time varying frequencies. The sinusoid used was:

$$T = \begin{cases} \frac{1}{2} \left[\sin \frac{2\pi}{50} t + \sin \frac{2\pi}{40} t \right]; & t = 0 \text{ to } 299 \\ \sin \frac{2\pi}{25} t; & t = 299 \text{ to } 400 \\ \sin \frac{2\pi}{15} t; & t = 400 \text{ to } 600 \\ \sin \frac{2\pi}{30} t; & t = 600 \text{ to } 800 \end{cases}$$

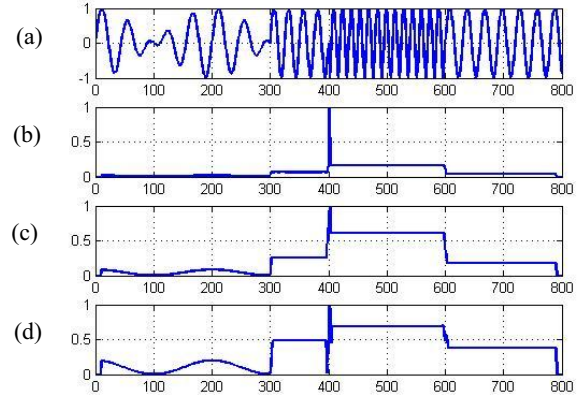


Figure 1: *Experimental results for a sinusoid signal of varying and mix frequencies.*

It is evident from Figure 1 that VTEO and S-VTEO captures the running estimate of signal energy with better temporal resolution than TEO, i.e., VTEO and S-VTEOs successfully track down the smallest variations in frequency as well as amplitude with significant changes in its amplitude.

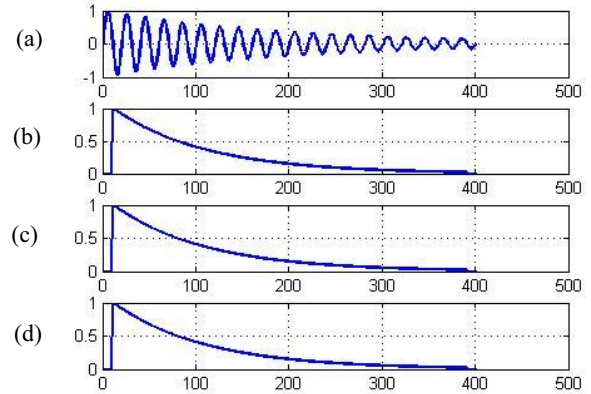


Figure 2: *Results for an exponentially damped sinusoid.*

Figure 2 and Figure 3 shows the TEO, VTEO and S-VTEO output for an exponentially damped sinusoid and a chirp signal, respectively. In Figures 4, 5 and 6, the output of the operators on the real speech data is compared. Figure 4, 5, and 6, respectively, shows the responses of the operators for a female speech, male speech and an infant cry signal.

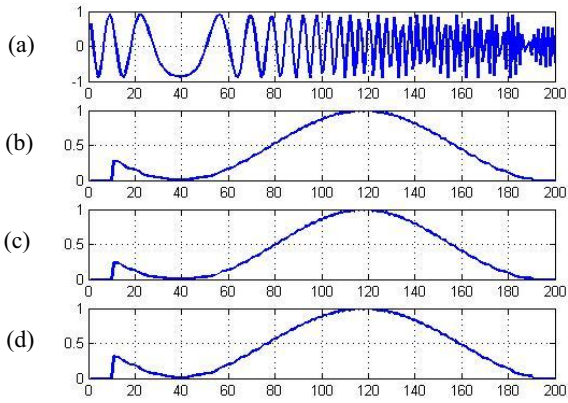


Figure 3: Results for a chirp signal.

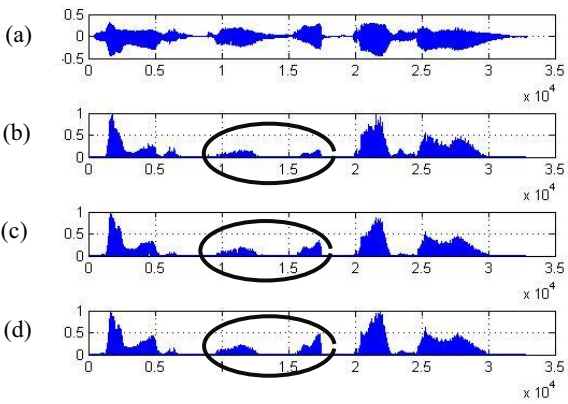


Figure 4: Results for a female speech signal

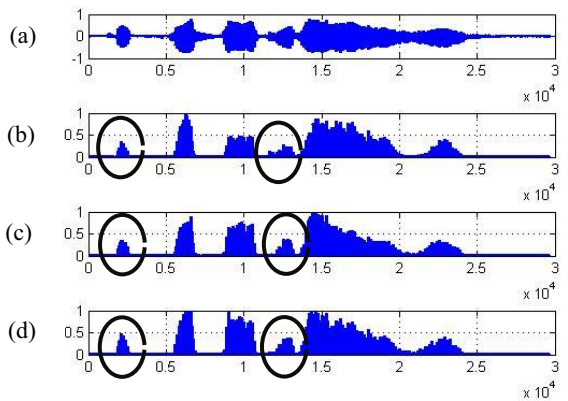


Figure 5: Results for a male speech signal

It is evident from the figures that all the operators perform equally well, in determining the running estimate of the energy, in the case of simple signals like synthetic signals in Figure 2 and 3 because there is significant difference in the consecutive samples of the signal. These results show that the VTEO and S-VTEO output is similar to that of the traditional TEO and therefore prove the validity of the proposed operators which generalizes TEO.

However, for signals like shown in Figures 1, 4, 5 & 6 where this difference is not so significant TEO fails to track

the energy fluctuations whereas it is clearly reflected in the proposed VTEO and S-VTEO output. Regions highlighted in Figures 4, 5, & 6 show that the proposed operator brings out the hidden dependencies and dynamics in a particular speech/infant cry segment.

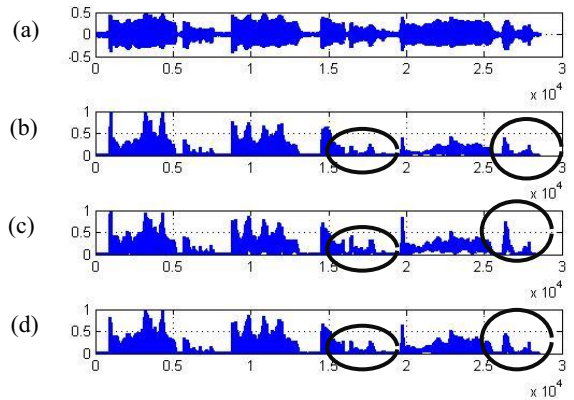


Figure 6: Results for an infant cry signal

5. Summary and Conclusions

In this paper, two new operators, viz. VTEO and S-VTEO, to track the running estimate of energy of a signal are developed by generalizing the concept of TEO. The effectiveness and validity of the proposed operators are demonstrated for various synthetic and real speech signals. The analyses provided in the paper demonstrate promising research directions to develop new feature extraction algorithms which can be used in speech/speaker recognition applications, for example, speaker recognition in the noisy environment.

6. Acknowledgements

The authors would like to thank DA-IICT authorities for their kind support and cooperation in carrying out this research work.

7. References

- [1] Teager, H. M., "Some observations on oral air flow during phonation", IEEE Trans. Acoust., Speech, Signal Processing, 28:599-601, 1980.
- [2] Zhau, G., Hansen, J. H. L., Kaiser, J. F., "Non-linear feature based classification of speech under stress," IEEE Trans. Speech Audio Processing, vol. 9, pp.201-216, 2001.
- [3] Kaiser, J. F., "On a simple algorithm to calculate 'energy' of a signal", in Proc. Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP'90, 1:381-384, Apr 1990.
- [4] Patil, H. A., Speaker recognition in Indian languages: A feature based approach, Ph.D. Thesis, Department of Electrical Engineering, IIT Kharagpur, Jul 2005.
- [5] Maragos, P., Quatieri, T., and Kaiser, J. F., "On amplitude and frequency demodulation using energy operators", IEEE Trans. Signal Processing, 4:1532-1550, Apr 1993.
- [6] Kaiser, J. F., "Some useful properties of Teager's energy operator," in Proc. Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP'93, 3:149-152, 1993.
- [7] Quatieri, F. T., "Discrete time speech signal processing", Pearson education, 570-576, 2004.
- [8] Jabloun F. and Cetin A. E., "Teager energy based feature parameters for speech recognition in car noise", IEEE Signal Processing Letters, 6(10), Oct 1999.
- [9] Sinder, D. J., Speech Synthesis using an Aerofrictive Model, Ph.D Thesis, Rutgers University, New Brunswick, NJ, Oct 1999.