A MULTIBAND EXPONENTIAL RATE OPERATOR FOR MUSICAL TRANSIENT ANALYSIS

Ramamurthy Mani and S. Hamid Nawab

ECS Department, Boston University

ABSTRACT

The *Exponential Rate Operator (ERO)* is presented for determining the "instantaneous exponential rate" of the amplitude modulation during musical transients. Its extension to multiband signal representations such as STFT and wavelet transforms is also described. Sensitivity of the *ERO* to white noise is examined and computational efficiency of the *STFT*-based *ERO* is discussed. Examples involving synthetic and real musical transients illustrate the usefulness of *ERO* analysis.

1. EXPONENTIAL RATE OPERATOR

A number of signal models [1, 2, 3] have been developed to analyze signals with time-varying amplitude and frequency modulations. A signal model suitable for the analysis of acoustic transients is :

$$x(t) = A \exp\left\{ \int_0^t (s(\tau) + jw(\tau))\, d\tau \right\}, \qquad (1)$$

where $s(t)$ and $w(t)$ are real-valued functions of t. In particular, we refer to $s(t)$ as representing the "instantaneous exponential rate"[1] of $x(t)$. The function $w(t)$ is called the "instantaneous frequency" of $x(t)$.

From (1), $s(t)$ can be expressed as :

$$s(t) = \frac{d}{dt}\left(\ln|x(t)|\right) \qquad (2)$$

We can thus view $s(t)$ as being obtained by applying an *exponential rate operator (ERO)* $g\{.\}$ to the signal $x(t)$. That is,

$$g\{x(t)\} = \frac{d}{dt}\left(\ln|x(t)|\right) \qquad (3)$$

We may also express (2) as

$$s(t) = \frac{\ln|x(t)| - \ln|x(t-\Delta)|}{\Delta} = \frac{1}{\Delta}\ln\left(\frac{|x(t)|}{|x(t-\Delta)|}\right), \qquad (4)$$

where Δ is a small decrement in time.

[1] For an exponential signal of the form $x(t) = e^{at}$, we call a the "exponential rate". The terms "attack-rate" and "decay-rate" are also associated with this parameter. The parameter $\frac{1}{a}$ is called the "time-constant" of the signal.

2. CONTINUOUS-TIME ERO FOR MULTIBAND ANALYSIS

Multiband signal representations have been used to capture amplitude and frequency modulations in different frequency bands [4]. Similarly, the rate operator can be extended to track exponential rates in different spectral bands by including a bandpass filter in its definition. Such an inclusion is justified because bandpass filters preserve purely exponential amplitude modulations.

The multiband exponential rate operator is defined as

$$g_\Omega\{x(t)\} = \frac{d}{dt}\left(\ln|x_\Omega(t)|\right) \qquad (5)$$

Here, $x_\Omega(t)$ is the output of a bandpass filter with center frequency Ω, when its input is $x(t)$. Such bandpass filtering is implicitly performed within well-known time-frequency representations such as the *STFT* and the wavelet transform.

The rate operator output has values in the range $(-\infty, \infty)$. For convenience, we impose a normalization on the rate operator that performs a bijective mapping of values in $(-\infty, \infty)$ to values in $(-1, 1)$, preserving symmetry about 0. One of the functions that provides such a mapping is the hyperbolic tangent function. The normalized multiband *ERO* can be defined as

$$G_\Omega\{x(t)\} = \tanh\left[g_\Omega\{x(t)\}\right] = \tanh\left[\frac{d}{dt}\left(\ln|x_\Omega(t)|\right)\right] \qquad (6)$$

Decay transients have normalized instantaneous exponential rate values in the range $(-1, 0]$, while attack transients have values in the range $[0, 1)$. Furthermore, it should be noted that the normalized exponential rate preserves the property of the exponential rate in that, changing its sign is equivalent to swapping attacks and decays with the same time constant.

3. DISCRETE-TIME ERO

For digital musical processing we have formulated a discrete-time approximation to our continuous-time rate operator. This approximation is obtained from (4) by replacing the time-decrement Δ with NT, where T is the sampling period and N is a positive integer. The discrete-time *ERO* is given by :

$$g\{x[n]\} = \frac{1}{NT}\ln\left(\frac{|x[n]|}{|x[n-N]|}\right), \qquad (7)$$

0-7803-2431-5/95 $4.00

Extending this operator for multiband analysis and applying the normalization of (6) gives

$$G_\omega\{x[n]\} = \tanh\left(\frac{NT}{2} g_\omega\{x[n]\}\right) = \frac{|x_\omega[n]| - |x_\omega[n-N]|}{|x_\omega[n]| + |x_\omega[n-N]|} \quad (8)$$

Here, $x_\omega[n]$ is the output of a bandpass filter with center frequency ω, when its input is $x[n]$. Once again, such bandpass filtering is implicitly performed by the discrete-time STFT and wavelet transforms.

4. EFFECT OF WHITE NOISE ON THE ERO

Let us consider a signal of the form $x[n] = e^{an}e^{j\omega_0 n} + w_1[n] + jw_2[n]$, where $w_1[n]$ and $w_2[n]$ are stationary zero-mean independent Gaussian white noise processes with variance σ^2. We will now determine the mean and variance of the normalized exponential rate of the signal $x[n]$. In the following derivations, we assume that the impulse response of the bandpass filter was obtained by modulating a rectangular window with a complex exponential of frequency ω_0.

Equation (8) may also be written as

$$G_{\omega_0}\{x[n]\} = f(X_1[n], X_2[n]) = \frac{\sqrt{X_1[n]} - \sqrt{X_2[n]}}{\sqrt{X_1[n]} + \sqrt{X_2[n]}}, \quad (9)$$

where $X_1[n] = |x_{\omega_0}[n]|^2$ and $X_2[n] = |x_{\omega_0}[n-N]|^2$. If we assume that the bandpass filter has an impulse response that is N_w points long, it can be shown that the mean and variance of the random processes $X_1[n]$ and $X_2[n]$ are given by

$$\mu_{X_1}[n] = 2\frac{\sigma^2}{N_w} + p^2[n], \quad \mu_{X_2}[n] = 2\frac{\sigma^2}{N_w} + p^2[n-N] \quad (10)$$

$$\begin{aligned}
\sigma^2_{X_1}[n] &= 4\frac{\sigma^2}{N_w}\left(\frac{\sigma^2}{N_w} + p^2[n]\right) \\
\sigma^2_{X_2}[n] &= 4\frac{\sigma^2}{N_w}\left(\frac{\sigma^2}{N_w} + p^2[n-N]\right) \quad (11) \\
\sigma^2_{X_1X_2}[n] &= \begin{cases} 0, & for\ N_w \le N \\ 4(\dot{p}[n]p[n-1])\frac{(N_w-N)\sigma^2}{N_w^2} + \\ \quad 4\frac{\sigma^4(N_w-N)^2}{N_w^4}, & for\ N_w > N \end{cases}
\end{aligned}$$

where $p[n]$ is the convolution of e^{an} with a rectangular window of length N_w.

The mean and variance of $G_{\omega_0}\{x[n]\}$ are estimated using the Taylor series approximation [5]. Using the second order approximation for the mean and the first order approximation for the variance, we obtain

$$\begin{aligned}
\mu_G[n] &= f(\mu_{X_1}[n], \mu_{X_2}[n]) + \frac{1}{2}\left(\frac{\partial^2 f}{\partial X_1^2}\right)\sigma^2_{X_1}[n] + \\
&\quad \frac{1}{2}\left(\frac{\partial^2 f}{\partial X_2^2}\right)\sigma^2_{X_2}[n] + \left(\frac{\partial^2 f}{\partial X_1 \partial X_2}\right)\sigma^2_{X_1X_2}[n] \\
\sigma^2_G[n] &= \left(\frac{\partial f}{\partial X_1}\right)^2\sigma^2_{X_1}[n] + \left(\frac{\partial f}{\partial X_2}\right)^2\sigma^2_{X_2}[n] + \\
&\quad 2\left(\frac{\partial f}{\partial X_1}\right)\left(\frac{\partial f}{\partial X_2}\right)\sigma^2_{X_1X_2}[n] \quad (12)
\end{aligned}$$

The expressions for the mean and variance were found to closely match experimental simulations.

The expressions in (12) were used to analyze the effect of N_w on the bias and variance of the exponential rate estimate. A signal of the form $x[n] = b^n e^{j\omega_0 n} + w_1[n] + jw_2[n]$ was considered. Here, $w_1[n]$ and $w_2[n]$ were zero-mean stationary independent Gaussian white noise random processes. Using (12), the bias and variance of the exponential rate estimate were plotted as functions of N_w for different values of b (see Figure 1). The plot shows that the exponential rate estimate has a larger bias and variance for smaller exponential rate signals ($b = 1.01$) as compared to larger exponential rate signals ($b = 1.04$). Therefore, greater emphasis has to be placed on the smaller exponential rates when N_w is being chosen for a particular application. Since the plots also show that smaller exponential rate signals show a decrease in both bias and variance with an increase in N_w, larger values of N_w are preferable.

We have also conducted a similar analysis on the effect of varying N while keeping N_w constant. Since the exponential rate estimates of signals with smaller exponential rates tend to have a larger bias and variance, the choice of N is also dependent on these rates. We have found that for smaller exponential rate signals, an increase in N results in a reduced bias but an increased variance. Therefore, the value of N has to be chosen such that a favourable tradeoff is obtained between bias and variance.

Finally, we have also demonstrated through simulations that a limited amount of averaging performed on $X_1[n]$ and $X_2[n]$ reduces the variance in the exponential rate estimate at the expense of a slight increase in bias.

5. EFFICIENT STFT-BASED ERO

In this section, we discuss an efficient method for implementing an *STFT*-based *ERO*. The need for such efficiency is particularly acute because (as we shall discuss below) for many practical applications, the *ERO*-associated *STFT* has to have a high degree of overlap between consecutive analysis window positions. Given such a need, the *STFT*-based *ERO* can take advantage of the recently proposed [6, 7] overlap-pruned *FFT* algorithm for *STFT* analysis with highly overlapped rectangular windows.

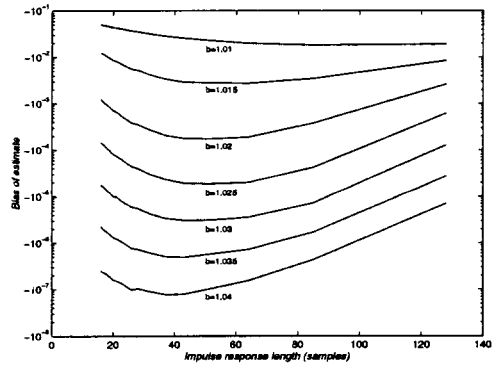
In (8) we assumed that the bandpass filter output $x_{\omega_0}[n]$ is sampled at the same rate as the signal $x[n]$. Therefore the time interval between two consecutive exponential rate estimates was the same as the sampling period of the original signal. If the attack or decay transients that occur in a particular application are of long duration (many times the duration of the bandpass filter's impulse response duration), then the exponential rate need not be determined at such frequent intervals of time. This in turn implies that the underlying *STFT* computation does not have to have significant overlap between consecutive window positions. However, musical transients can be as short as a few milliseconds, which is on the order of the duration of typical *STFT* analysis windows. We have found that in such situations *STFT*-based *ERO* analysis loses significant temporal resolution unless there is substantial (90% or more) overlap between consecutive window positions. To demonstrate this fact, we conducted *STFT*-based *ERO* analysis

of the synthetic monochromatic transient signal shown in Figure 2. Based on an exhaustive set of *STFT*-based *ERO* analyses using all possible window overlaps and window-to-signal alignments, we obtained the data displayed in Figure 3. For each window-overlap percentage, we have shown the *worst case* performance from among the *ERO* analyses with different window-to-signal alignments. Since each *ERO* analysis usually produced several exponential rate estimates (within the duration of the exponential transient), its performance is characterized in terms of the error corresponding to the best estimate produced by it. It is clear from Figure 3 that greater window overlaps result in better performance.

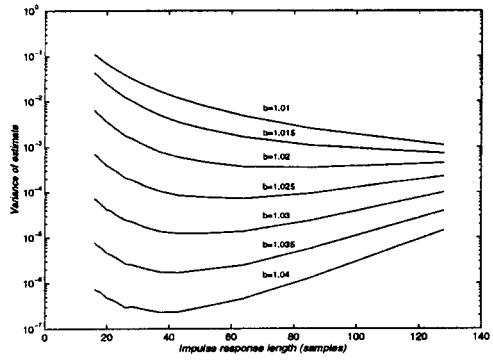
6. EXAMPLES

Figure 4 shows the results of processing a synthetic signal with the multiband discrete-time *ERO*. The signal is a sum of two sinusoidal components whose frequencies are $2kHz$ and $4kHz$. Both components have three different temporal regions - an attack, a steady part and a decay. Figure 4(a) shows the location of the spectral peaks in the signal. Figures 4(b) and 4(c) show the rate operator output for the $2kHz$ and $4kHz$ components respectively. The *ERO* output clearly shows the three different temporal regions for each component. The values of the instantaneous exponential rate were used to determine the time constants of the attacks and decays. These values were found to be good approximations of the parameters that were used to generate the signal.

Figure 5 shows the results of processing a real music signal using the multiband exponential rate operator. The music signal corresponds to an English-horn playing a single note. The *ERO* was used to track the instantaneous exponential rate in the signal's three highest spectral components (which are at $1kHz$, $2kHz$ and $3kHz$). A threshold on the *ERO* output was used to detect transients whose exponential rates were greater than $25Hz$ (time constants less that 40m-sec). This was used as a basis for calculating the duration of the attack and decay transients. Figure 5(a) shows the results of this calculation for the English horn note. The three highest spectral components are shown as streaks of black and white. The black regions indicate the attack and decay transient regions. The white regions correspond to the steady part of the note. Figure 5(b) shows the *ERO* output for the attack and decay regions of the first spectral component which lies at 1kHz.



(a)



(b)

Figure 1: (a) Plot of bias versus N_w (b) Plot of variance versus N_w. Bias and variance are plotted on a logarithmic scale. The underlying signal was of the form $x[n] = b^n e^{j\omega_0 n} + w_1[n] + jw_2[n]$. Here, $w_1[n]$ and $w_2[n]$ are independent zero-mean Gaussian white noise random processes with. The variance of both $w_1[n]$ and $w_2[n]$ was 1000. N was always chosen to be equal to N_w.

7. REFERENCES

[1] M. D. Freedman, "Analysis of Musical Instrument Tones," *J. Acoust. Soc. Am.*, vol.41, no. 4, pp. 793-806, Apr., 1967.

[2] R. J. McAulay, T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. Acoust. Speech and Sig. Proc.*, ASSP-34, no. 4, pp. 744-54, August, 1986.

[3] W. Strong, M. Clark, "Synthesis of Wind-Instrument Tones," J. Acoust. Soc. Am., vol. 42, no. 1, pp. 39-52, Jan., 1967.

[4] A. C. Bovik, T. F. Quatieri, "AM-FM Energy Detection and Separation in Noise Using Multiband Energy Operators," *IEEE Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3245-65, Dec., 1993.

[5] A. H-S. Ang, W. H. Tang, "Probability Concepts in Engineering Planning and Design," vol.1, John Wiley and Sons, 1975.

[6] M. M. Covell, J. Richardson, "A New, Efficient Structure for the Short-Time Fourier Transform, with an Application in Code-Division Sonar Imaging", *Proceed. ICASSP 1991*, vol. 3, pp. 2041-44, Toronto, 1991.

[7] D. E. Paneras, R. Mani, S. H. Nawab, "STFT Computation Using Pruned FFT Algorithms," *IEEE Sig. Proc. Letters*, vol. 1, pp. 61-63, April, 1994.

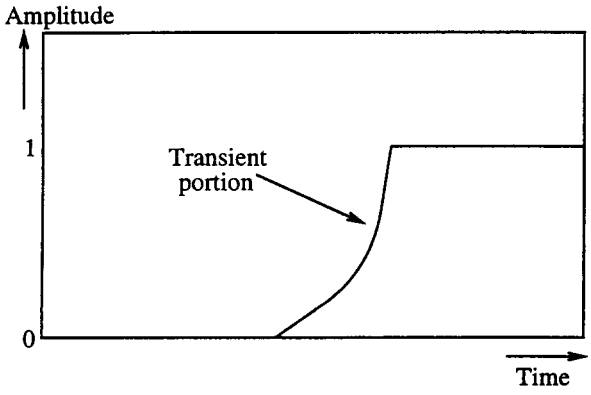


Figure 2: Synthetic attack transient used to demostrate the need for high degree of overlap between consecutive STFT window positions.

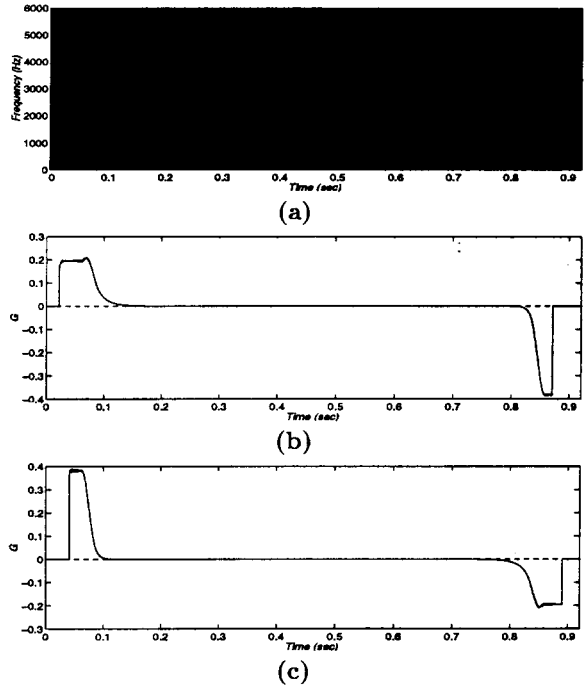


Figure 4: (a) Spectral peaks of synthetic signal (b) ERO output for 2kHz component (c) ERO output for 4kHz component.

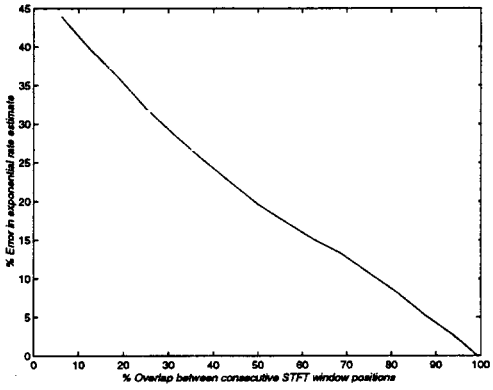


Figure 3: Plot of % error in exponential rate estimate versus % overlap between consecutive $STFT$ window positions. The underlying signal was of the form $x[n] = (1.02)^n e^{j\omega_0 n + j\phi}$. The ERO was applied to the $STFT$ channel corresponding to ω_0. A window length of 128 samples was used throughout this experiment.

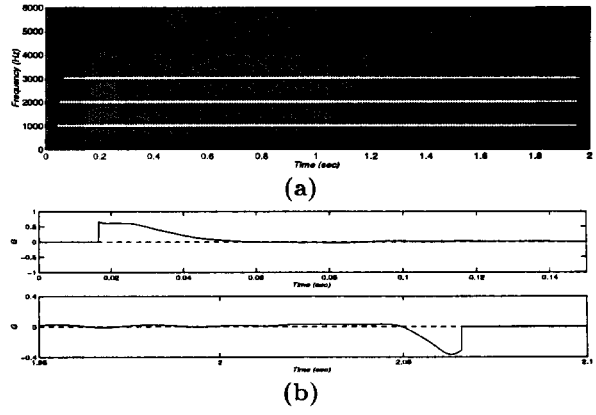


Figure 5: (a) The three highest spectral peaks of a English horn note - located at $1kHz$, $2kHz$ and $3kHz$. The black streaks indicate the attack and decay regions of each component. The white streaks indicate the steady part. (b) ERO output for attack and decay regions of the $1kHz$ component.