ANALYSIS AND RESYNTHESIS OF MUSICAL INSTRUMENT SOUNDS USING ENERGY SEPARATION

Robert B. Sussman and Mark Kahrs

CAIP Center
Rutgers University, P.O. Box 1390
Piscataway, NJ 08855-1390
rsussman@caip.rutgers.edu, rbs@ade.att.com
kahrs@caip.rutgers.edu

ABSTRACT

Energy separation algorithms using Teager's energy operator have been used to analyze modulations in speech resonances, and energy separation in bird song. Here we apply energy separation and the energy operator to the analysis and resynthesis of musical instrument sounds. We show how the energy operator can be used to (a) examine vibrato and tremolo (b) precisely determine frequencies of harmonics (c) determine synthesis parameters for an excitation/filter model.

1. INTRODUCTION

Newer techniques to synthesize musical sound use physical models to represent the instrument. A mathematical model of the instrument is devised with control parameters that specify the sound. These parameters can often be determined through analysis of a recording of the instrument being modeled.

One common class of models is that of an excitation signal feeding a resonant filter. For example, in a guitar sound the excitation signal could represent the physical excitation, or pluck, and the resonant filter could represent the actual guitar string. In this example the control parameters are the excitation signal, and the filter poles and zeros.

We are interested in two aspects of musical instrument modeling: 1) separation of the excitation and resonator 2) modeling of performance parameters such as tremolo (AM) or vibrato (FM). We will show how "Energy Separation" can be used to determine synthesis parameters for an excitation/filter model, and also be used to follow tremolo and vibrato.

Mark Kahrs was partially supported by the ARPA tri-service MHDL effort, monitored by ARL DAAL01-93-K-3370.

2. ENERGY SEPARATION

Energy separation algorithms use the Teager energy operator to separate the amplitude modulated (AM) and frequency modulated (FM) components of an AM-FM signal. The Teager energy operator is a very simple time domain operator. Many properties of this simple operator were elucidated by Kaiser [1]. Examining a few of these properties offers a more intuitive understanding of energy separation algorithms.

Kaiser showed how the energy operator for continuous time signals,

$$\Psi_c[x(t)] = \dot{x}^2(t) - x(t)\ddot{x}(t) \tag{1}$$

can be used to find the amplitude and frequency of a signal of the form, $s(t) = A\cos(\omega_c t + \theta)$ by applying the energy operator to both $s(t)$ and $\dot{s}(t)$. It is easy to show that, $\Psi_c[s(t)] = A^2\omega_c^2$. Also, $\Psi_c[\dot{s}(t)] = A^2\omega_c^4$. Therefore, the instantaneous amplitude and frequency can be determined by:

$$\omega_c = \frac{1}{2\pi}\sqrt{\frac{\Psi_c[\dot{s}(t)]}{\Psi_c[s(t)]}} \tag{2}$$

$$\mid A \mid = \frac{\Psi_c[s(t)]}{\sqrt{\Psi_c[\dot{s}(t)]}} \tag{3}$$

Kaiser showed similar properties of the discrete-time energy operator:

$$\Psi_d[x(n)] = x^2(n) - x(n+1)x(n-1) \tag{4}$$

Which can be used to find the amplitude and frequency of a signal of the form, $x(n) = A\cos(\Omega_c n + \theta)$, by applying the energy operator to both $x(n)$ and the first difference, $y(n) = x(n) - x(n-1)$. Kaiser showed that, $\Psi_d[x(n)] = A^2\sin^2(\Omega_c)$, and also, $\Psi_d[y(n)] = 4A^2\sin^2(\Omega_c/2)\sin^2(\Omega_c)$. Therefore, the discrete time

0-7803-3192-3/96 $5.00©1996 IEEE

versions of (2) and (3) are:

$$\Omega_c = \arccos\left(1 - \frac{\Psi[y(n)]}{2\Psi[x(n)]}\right) \quad (5)$$

$$|A| = \sqrt{\frac{\Psi[x(n)]}{\sin^2(\Omega_c)}} \quad (6)$$

Maragos, Kaiser and Quatieri [2] show how to use the energy operator to estimate the instantaneous amplitude and instantaneous frequency of an AM-FM signal. Given the AM-FM signal x(n):

$$x(n) = a(n)\cos\phi(n) \quad (7)$$

$$\phi(n) = \Omega_c n + \Omega_m \int_0^n q(m)dm + \theta \quad (8)$$

with instantaneous frequency defined in terms of carrier frequency Ω_c, and maximum frequency deviation Ω_m:

$$\Omega_i(n) = \frac{d}{dn}\phi(n) = \Omega_c + \Omega_m q(n) \quad (9)$$

Maragos, et al. define an energy separation algorithm called "DESA-1":

$$\Omega_i(n) = \arccos\left(1 - \frac{\Psi[y(n)] + \Psi[y(n+1)]}{4\Psi[x(n)]}\right) \quad (10)$$

$$|a(n)| = \sqrt{\frac{\Psi[x(n)]}{1 - \left(1 - \frac{\Psi[y(n)]+\Psi[y(n+1)]}{4\Psi[x(n)]}\right)^2}} \quad (11)$$

$$y(n) = x(n) - x(n-1) \quad (12)$$

These estimates are valid under the assumptions that the bandwidths of $a(n)$, $\Omega_i(n)$, and Ω_m are small compared to Ω_c.

Maragos, et al. used energy separation to analyze modulations in speech resonances. They used a Gabor bandpass filter centered at an estimate of the resonance frequency to filter out a speech resonance. The output of that filter was then input to the DESA-1 algorithm. In this work we use a similar technique for the analysis of musical instrument sounds.

3. ANALYSIS

Musical sounds generated by resonate structures have spectra composed of harmonics spaced very close to integer multiples of a fundamental frequency. Our analysis consisted of first estimating the fundamental frequency of the sound. We did this through peak picking of the discrete-time Fourier transform (DTFT). Next a bank of bandpass filters was created to separate the harmonics. The output of each bank was then passed into the DESA-1 algorithm to determine the AM and FM components of each harmonic.

Like Maragos, et al. we used Gabor bandpass filters given by:

$$h(n) = e^{-(bn)^2}\cos(2\pi f_c nT) \quad -N \leq n \leq N \quad (13)$$

Where f_c is the center frequency, T is the sampling period, bandwidth is chosen by setting b to BW$\cdot T\sqrt{2\pi}$, and N is chosen so that h(n) is small at the edges, that is $exp(-(bn)^2) = 10^{-5}$. The center frequency of the first filter was set to the estimate of the fundamental frequency. Each additional filter in the bank had its center frequency set to an integer multiple of the fundamental frequency. This was a logical choice that worked well with the sounds that we analyzed. The choice for bandwidth of the filter was not obvious. The filter needs to be wide enough to prevent filtering out the modulations that we are trying to detect. However, if the filter is too wide, adjacent harmonics will interfere. We achieved good results on many sounds using a bandwidth of one quarter the fundamental frequency.

The DESA-1 algorithm was applied to a plucked string (electric guitar), and a plucked string with vibrato (electric guitar).

The DESA-1 algorithm was applied to the first 400 ms of this guitar pluck which was recorded at a sampling frequency of 22050 samples per second. Figure 1 shows the results for the first harmonic. Note after the initial transient the relatively constant frequency and decaying amplitude. Similar results were obtained for higher harmonics.

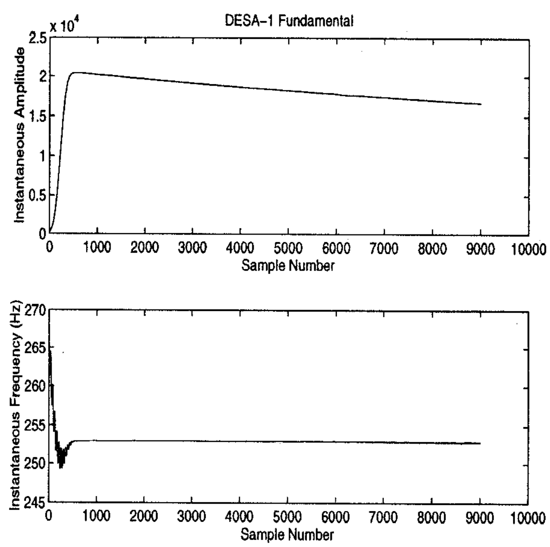

Figure 1: DESA-1 applied to guitar pluck.

A good example of how the DESA-1 algorithm can find vibrato and tremolo is found in the third harmonic of a guitar pluck played with vibrato recorded at 22050 samples per second. This is shown in figure 2.

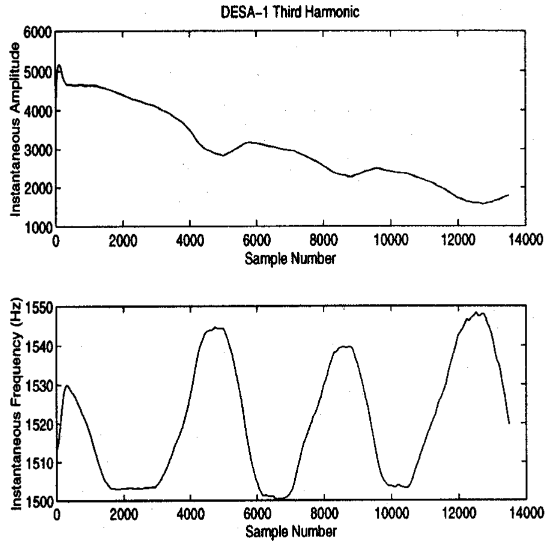

Figure 2: FM and AM modulations of guitar pluck with vibrato.

The use of vibrato is clearly found on this example. It is interesting to look at the relationship between vibrato and tremolo. Looking at the frequency track, one sees periods of relatively flat frequency followed by a pulse of increased frequency as the string is stretched. After the string is stretched, the amplitude track bumps up slightly due to the energy added to the string from the vibrating finger.

4. SEPARATION OF EXCITATION AND RESONATOR

An appropriate model for musical instruments with strings that are plucked or struck is a short excitation signal fed into a bank of resonators[3]. We model each harmonic of the sound with a single resonance. A filter is designed to implement these resonators. The poles are located to give the proper resonance frequencies, and the proper decay rate for each resonance. Once the filter is constructed, it is inverted, and used to extract an excitation signal that can be used to resynthesize the instrument sound when passed through the forward filter. We define the transfer function of the resonant filter as:

$$H(z) = \frac{B(z)}{A(z)} \tag{14}$$

$$A(z) = \prod_{k=1}^{N}(1 - d_k z^{-1}) \tag{15}$$

$$B(z) = \prod_{k=1}^{M}(1 - c_k z^{-1}) \tag{16}$$

The resonances of the filter are determined by its poles. The zeros will affect the excitation signal. The filter $H(z)$ must be stable, and must have a real impulse response. Therefore, the poles must come in conjugate pairs, and be located inside the unit circle. To estimate the pole locations we used energy separation to determine instantaneous frequency tracks, $\Omega_i(n)$ and amplitude functions $a(n)$ for each partial. Next, we manually picked a steady-state portion of the signal $R_1 \leq n \leq R_2$ by visually inspecting the signal waveform. Next we estimated the angle and magnitude of each pole as follows:

$$\angle c_k = \frac{1}{R_2 - R_1 + 1} \sum_{n=R_1}^{R_2} (\Omega_i(n)) \tag{17}$$

$$| c_k | = \frac{1}{R_2 - R_1 + 1} \sum_{n=R_1}^{R_2} (a(n+1)/a(n)) \tag{18}$$

Hanson, Maragos, and Potamianos [4] used energy separation with an iterative approach to find speech formant frequencies. They used a similar approach. The estimate of the formant frequency was the average of the output of the instantaneous frequency portion of the energy separation algorithm.

Now that we have determined the pole locations we need to pick a suitable numerator polynomial $B(z)$, and then find the excitation by filtering the original signal with the inverse filter. Laroche and Meillier[3] discuss the inverse filtering problem as an ill-conditioned problem. If the resonant filter has deep valleys in its magnitude response, the corresponding inverse filter will have large peaks. This could cause noise in the instrument sound to be amplified at these frequencies and swamp the resulting excitation signal. For example, all-pole filters generally exhibit high attenuation in the high frequency region. This will cause the inverse filter to have a large high frequency gain, and the excitation will be swamped with high frequency components that are not necessarily meaningful. Laroche and Meillier suggest the use of a parallel second-order cosine section filter because it is better conditioned. Also, $B(z)$, the numerator polynomial for this filter, is known to be minimum phase and therefore invertable. The numerator polynomial for this filter is:

$$B(z) = \sum_{i=1}^{p} \prod_{j \neq i} (1 - z_j z^{-1}) \tag{19}$$

We had good results using this method for the first six harmonics but larger filters became unstable due to the poor conditioning of the filter coefficients.

Figure 3 shows the excitation obtained from inverse filtering of a guitar pluck. Note that it is a short pulse followed by high frequency components left over because we only used the first six harmonics.

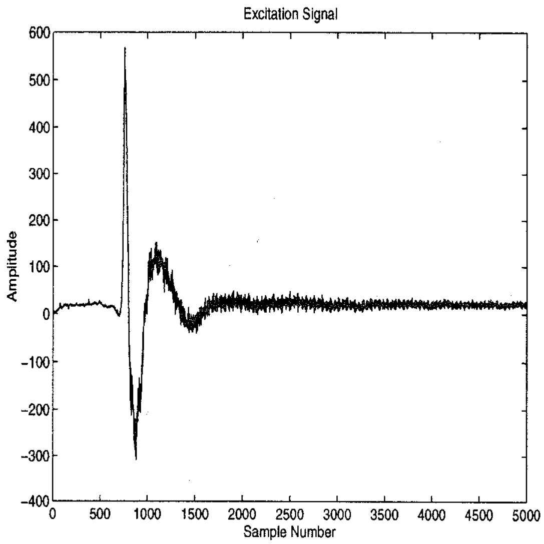

Figure 3: Excitation obtained from inverse filtering guitar pluck.

Resynthesis with this excitation achieves a close but not exact match. This is because of the limited number of poles permitted before the problems of numerical stability appear. Figure 4 shows a short segment from the middle of the original guitar signal and the corresponding segment from the middle of the resynthesized signal. Note that they are very similar, but not identical.

5. CONCLUSION

Discrete energy separation algorithms are useful for analyzing musical instrument sounds. We have shown the use of energy separation in the analysis of an electric guitar sound. Tremolo and vibrato was followed, and frequencies and decay rates of resonances in a resonant filter were found. We used this resonant filter to resynthesize the guitar sound using an excitation/filter model.

Another area to be investigated is the use of the output of the energy separation algorithm to find control parameters for the synthesis of musical instrument sounds. For example, the guitar vibrato and tremolo identified earlier in this paper could be used to control the resynthesis to produce a natural sounding vibrato. This is currently being worked on, along with a similar application of energy separation to piano sounds in [5].

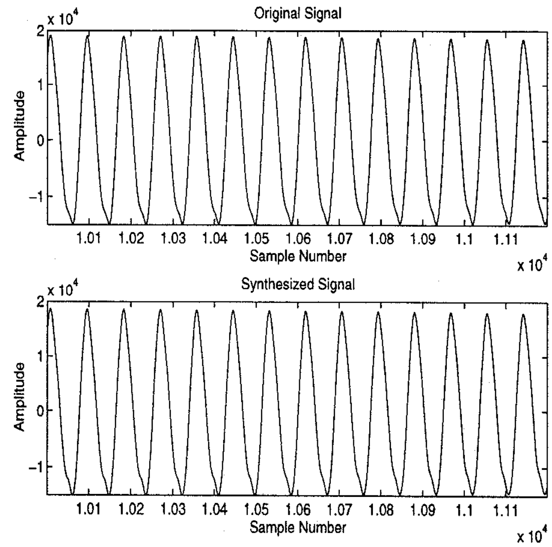

Figure 4: Original and resynthesis of portion of a guitar pluck.

6. REFERENCES

[1] J.F. Kaiser. On a simple algorithm to calculate the 'energy' of a signal. In *Proceedings of ICASSP*, pages 381–384, April 1990.

[2] Petros Maragos, James F. Kaiser, and Thomas F. Quatieri. Energy separation in signal modulations with application to speech analysis. *IEEE Transactions on Signal Processing*, 41(10):3024–3051, October 1993.

[3] Jean Laroche and Jean-Louis Meillier. Multichannel excitation/filter modeling of percussive sounds with application to the piano. *IEEE Transactions on Speech and Audio Processing*, 2(2):329–343, April 1994.

[4] Helen M. Hanson, Petros Maragos, and Alexandros Potamianos. A system for finding speech formants and modulations via energy separation. *IEEE Transactions on Speech and Audio Processing*, 2(3):436–443, July 1994.

[5] Robert B. Sussman. Analysis and resynthesis of musical instrument sounds using energy separation. Master's thesis, Rutgers, The State University of New Jersey, Graduate School, New Brunswick, In Progress.