Reproduction and Compression of Violin Tones through Spectral Interpolation Synthesis

Chris Niessl
Advisor: Roger Dannenberg

Undergraduate Research Senior Thesis
Carnegie Mellon University
April 30th, 2010

Abstract:

Synthesizing instrument performances have a variety of useful applications, from providing a computerized 'session musician' for improvisation and composition, to a form of source based compression for audio data. However, synthesizing instruments accurately is a challenging problem; merely splicing recorded samples for pitches ignores variation in timbre and dynamics that are affected by the style of play. In addition, using signal processing techniques such as spectrum analysis can generate large amounts of data. In this paper, we explore several methods of expanding Spectral Interpolation Synthesis in order to successfully synthesize violin tones.

Introduction:

In 1998, Spectral Interpolation Synthesis was shown to be successful in replicating trumpet notes played in succession by successfully capturing greater features of the sounds produced by the instrument than prior synthesis models. [1] The primary idea behind SIS was through extracting amplitude data of the signal into harmonic components of the instrument. By mapping that data into a series of envelopes, and having those envelopes normalized against the RMS of the signal, synthesis could be achieved without needing an expensive amount of computation or resources.

However, initial attempts at reducing the control data rate for applying SIS on Violin tones resulted in audible "smearing" of the tone. One significant reason, discovered by Mellody and Wakefield is that amplitude modulation is present as a result of frequency modulation from vibrato and finger motions on the strings [2]. In addition, whereas the sound from brass and woodwind instruments are produced can be modeled as simple resonators, string instruments introduce complications from Helmholtz motion interactions between the violin string and resin on the bow. The addition of possible time varying properties increases the difficulty of using spectral interpolation to detect those variations. [3] Thus, we explore adding a noise model to traditional synthesis in order to improve the quality of synthesis.

Methodology and Collection:

Initial testing was conducted on a single violin tone, a middle G pitch (G4) with a fundamental harmonic frequency of approximately 202hz. Phase Vocoder Analysis was used through the SNDAN software package [4] to obtain amplitude and frequency information for all multiple of the fundamental harmonic (202hz, 404hz …) up to the 22050hz, above the Nyquist rate for human audio perception. This resulted in a total of 1186 data points for 109 harmonics over 3 seconds of the sound. This information was then used to re-synthesize the original tone, through SIS, and used as a reference point of comparison against lower data rate signals.

Lower data rate signals were then obtained by encoding all the amplitude points for a harmonic as a new signal, and then re-sampling that signal at approximately 30hz. Each re-sampled harmonics signal was then recombined to form a control signal file. Spectral Interpolation Synthesis was then used to reconstruct the violin sound from the control signal, and compared against the reference point.

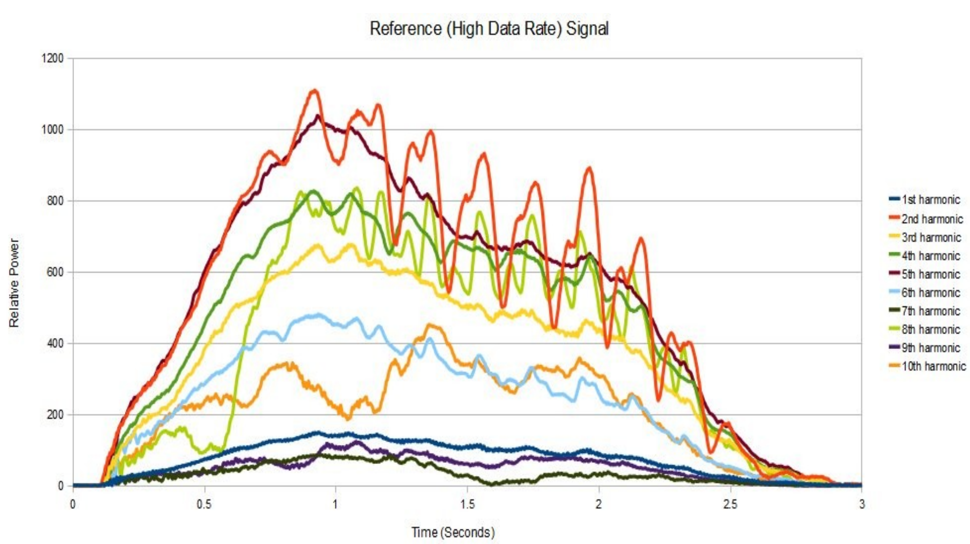
While SIS produced a very good approximation of sound, some finer details from the original tone (Such as scratches due to interactions between bow resin and string), were lacking in the tone due to the smoothing caused by re-sampling the sound at a lower rate. In order to compensate for the lost sound, additive synthesis was used to combine the tone reproduced via the control signal file with another tone produced through modeling noise in the original tone.

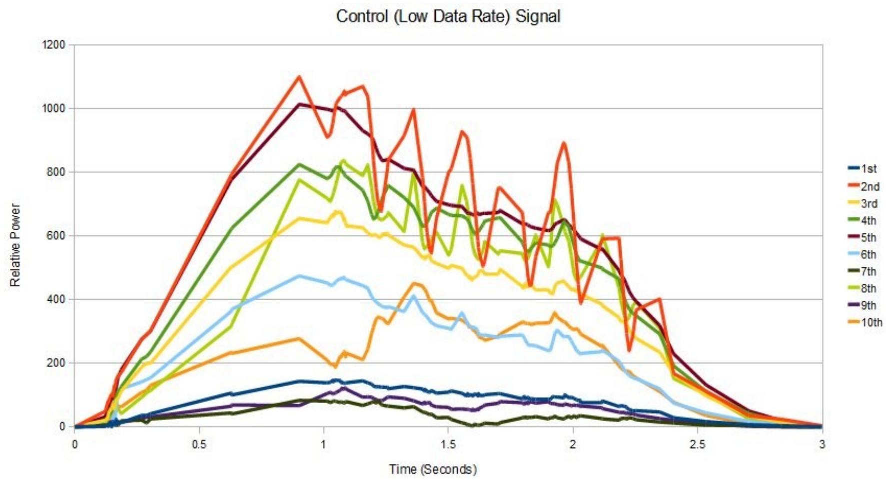
The Noise model was created by analyzing differences between the control signal and original tone for several pitches. Noise for each pitch was analyzed by capturing the differences in amplitude data points for each harmonic between the reference signal and the control signal. Spectral analysis was then taken on each harmonic via FFT in order to determine if the noise signal had frequency dependent properties. Once the noise model was completed, samples in the lower data rate signal were then reconstructed with the noise model to generate a combined tone.

Once the initial tone was verified to synthesize properly, several other tones were analyzed in order to extract envelope information and determine patterns. Envelopes between the high data rate and combined low rate tones were plotted to identify potential missing features in the tone. In addition, frequency variation was plotted against amplitudes for each harmonic, normalized to the RMS of the original signal.

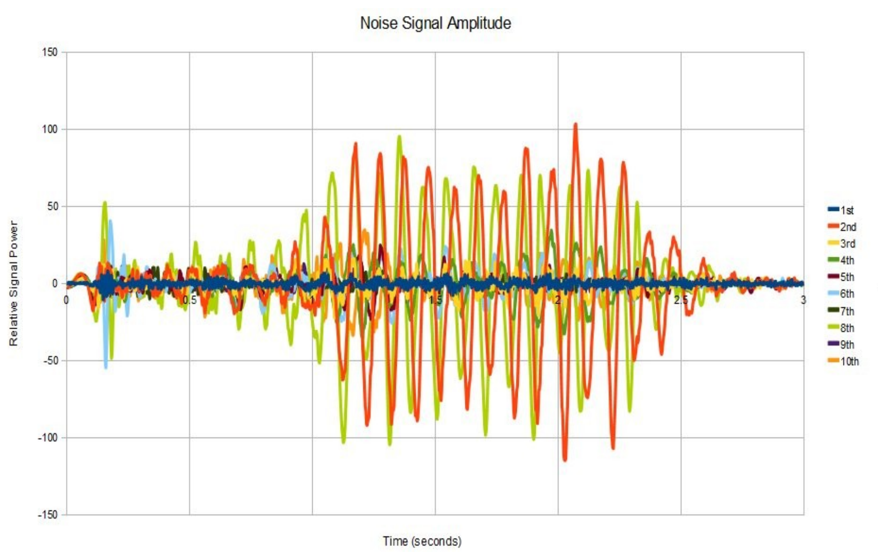
Data Analysis:

Below are two plots of the first 10 harmonics of each signal, with fundamental frequency of 202hz. The first graph consists of plotting all data points from the reference signal, captured at 404 data points per second (404hz) while the bottom graph consists of the low data-rate control signal plotted at 30hz. Relative power level refers to the amplitude used to reconstruct the tone through a 16-bit wav file. A maximum amplitude wave of 65,536 would result in an approximate 96dB difference to a signal with maximum amplitude of 1.0.

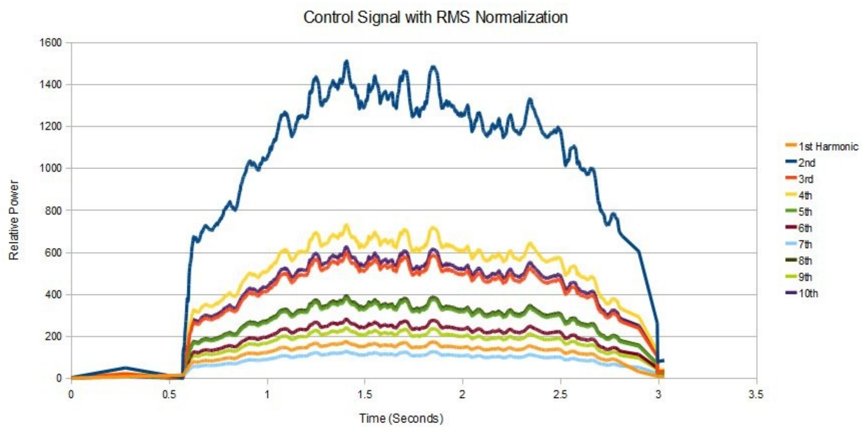




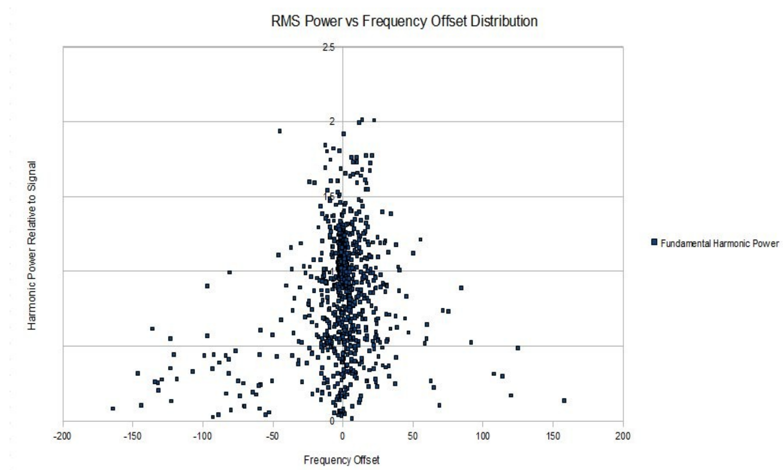
When comparing the reference synthesized tone against the original tone, the two are virtually indistinguishable, suggesting that SIS can reproduce the sound, given a sufficiently high data-rate (around 400 samples per second). While listening to tone synthesized through the low data-rate control signal without noise, the tone resembles a violin, however differences become a bit more noticeable when compared against the other two tones. This is consistent with Horner and Beauchamp's findings, which suggest that experienced musician detection of alterations are noticeable outside 16%[5]. A visual comparison of the two graphs reveals the result of reducing the sample rate. The following graph models the noise that was extracted and re-synthesized into the low-data rate control signal. A relative signal power of 100 corresponds to approximately a 20dB difference in volume level.



As you can see, several of the harmonics (the 2nd and 8th most prominently) display periodic variation from the control signal. When the noise is re-synthesized through SIS by modifying the original low-rate control signal with modifications from the signal above, a richer tone is achieved. Further analysis of more of the tones however, reveals that this model is short of being able to describe all possible violin tunes. An attempt to consolidate all information for all the harmonics into a single envelope with weights for each harmonic reveals the following plot:



Contrasting this plot against the previous ones for the synthesized tone, we notice that the distinct shapes for each of the harmonics is destroyed, and the sound reproduced by synthesizing those envelopes reflects that. Plotting each harmonics amplitude relative to the RMS of the signal at that point for all points in each tone reveals the following pattern:



Significant clustering occurs at 0hz offset from the original signal. In this region, most points fall around 85 to 135% of the RMS power, while greater scattering occurs as a result of a higher frequency offset. This reinforces Mellody and Wakefield's study, as well as provide further insight as to developing an improved synthesis model, suggesting that differences in the synthesized and original signal can be compensated by accounting for the difference in the fundamental frequency, and frequency used for analysis at that frame.

Conclusions and Future Work:

Using the methods described here, a violin tone can be re-synthesized at approximately 30% of the original size of the signal without sacrificing quality by successfully reducing the information needed for the control signal for SIS. However, there is room for improvement, as new models that integrate frequency dependent control information have the potential to reduce information by correlating amplitude differences in each harmonic with the frequency of the frame analyzed. In addition, psycho-acoustic analysis of amplitudes may reveal that some harmonics may not need to be included in order to achieve audibly indistinguishable reproduction.

Future progress on creating a successful model for violin synthesis can be achieved by expanding this model with the aforementioned observations. Once a successful model has been developed, future applications of

Acknowledgments / Sources:

In addition to my thesis advisor for providing support, I'd like to thank Yuxiang Liu for providing assistance in setting up SNDAN on my machine for expediting the analysis of violin tones.

[1] Dannenberg, Pellerin, and Derenyi. ``A Study of Trumpet Envelopes," in Proceedings of the International Computer Music Conference. San Francisco: International Computer Music Association (1998) pp 57-61.

[2] , Gregory H. Wakefield , "The time-frequency characteristics of violin vibrato: Modal distribution analysis and synthesis" Journal Acoustic Society America Volume 107, Issue 1, pp. 598-611 (January 2000)

[3] Horner, A., Beauchamp, J., and So, R. 2009. "Detection of Time-Varying Harmonic Amplitude Alterations due to Spectral Interpolations between Musical Instrument Tones," Journal of the Acoustical Society of America, 125(1), 492-502.

[4] Rob Maher, James Beauchamp "SNDAN, programs for sound analysis, resynthesis, display and transformation" University of Illinois at Urbana-Champaign (UIUC). http://dream.cs.bath.ac.uk/software/sndan/sndan.html

[5] Horner, A., and Beauchamp, J. 2003. "Discrimination of Sustained Musical Instrument Sounds Resynthesized with Randomly Altered Spectra", in 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '03), Mohonk, NY, pp. 169-172.