Non-harmonic Sinusoidal Modeling Synthesis Using Short-time High-resolution Parameter Analysis

Ozgur Izmirli

Center for Arts and Technology
Department of Mathematics and Computer Science
Box 5318, Connecticut College
New London, CT 06320, USA
oizm@conncoll.edu

ABSTRACT

This paper describes a method to perform analysis and resynthesis of non-harmonic and possibly simple polyphonic musical signals. Sinusoidal parameter extraction is performed using the high-resolution matrix-pencil method. Instead of operating on the complete signal all at once, a short-time approach has been developed both to reduce the computational complexity and for the resynthesis to be capable of reproducing gradual frequency variations of partials. Partial tracking that utilizes a birth-death strategy is performed using the extracted frame based parameters. Resynthesis of signals with closely spaced beating partials are demonstrated.

1. INTRODUCTION

Many models proposed for sinusoidal modeling synthesis use the short-time Fourier Transform (STFT) or similar window based methods as a means of parameter extraction. In these models, the analysis phase utilizes an analysis window that determines the spectral resolution. In order for the sinusoidal parameters to reflect the changes in the signal within small time proximity, the window is chosen to be as short as possible. Naturally, a short window has less amplitude smearing compared to longer windows. Amplitude smearing can be audible in the form of pre-echo and smoothed attack during silence to signal transitions.

For monophonic sounds the trade-off between frequency resolution and time resolution is selected such that the frequency resolution is just sufficient to resolve the partials of a harmonic signal. Usually, the resolution is set to be equal to the minimum partial frequency spacing, which corresponds to the fundamental frequency in a harmonic tone. Hence, in order to use the STFT for non-harmonic signals, and equivalently, simple polyphonic signals, the frequency resolution has to be increased at the expense of increasing the smearing effects mentioned above. One solution to this problem has been proposed in [1] in the form of an adaptive Q transform for musical signals. The above discussion assumes that the partials need to be resolved in the spectral transform, solely from an analysis point of view. From a perceptual perspective, it is also possible to state that it is more important to capture the modal structure in the resynthesis whether the partials are resolved or not [2]. This paper outlines a solution to the former, that is, taking an analysis viewpoint, by making use of a localized high-resolution analysis method.

2. SHORT-TIME ANALYSIS AND SYNTHESIS

2.1. Analysis and short-time interpretation

For the high-resolution spectral analysis, the matrix-pencil method is used [3]. The method depends on a signal model that consists of a sum of real sinusoids with exponential envelopes. Initially, frequencies and damping factors of the sinusoids in each frame of the input are estimated. Next, using these parameters, the amplitudes and phases are estimated by time-domain least squares minimization. As much as the matrix-pencil method is useful for improved resolution parameter extraction, it has a serious limitation in that the input signal has to be fairly short (in the order of 1000-3000 samples). This is due to the computation time requirements to perform the analysis and makes it infeasible to be directly applied to sounds that are even a few seconds long. It is therefore necessary to develop a technique in which this method can be utilized locally in time. A method that combines the frequency resolution advantages of the matrix-pencil method with a partial tracking method is outlined below.

The matrix-pencil method serves as a short-time frequency analyzer and according to this interpretation the signal is divided up into short frames. Sinusoidal parameters are extracted for each frame and the local signal, $x_j[n]$, is synthesized. According to this short-time interpretation the signal model can be written as :

$$x_j[n] = \begin{cases} \sum_{k=1}^{S_j} a_{j,k} e^{-n\alpha_{j,k}} \cos(2\pi f_{j,k} n - \phi_{j,k}) & 0 \le n < N \\ 0 & otherwise \end{cases} \quad (1)$$

$$X[m] = \sum_{j=0}^{T-1} x_j[m - jN], \qquad 0 \le m < TN \quad (2)$$

where S_j is the number of sinusoids for frame j, N is the number of samples per frame and T is the total number of frames. For frame j, component k, a, f, α, ϕ are the amplitude, frequency, damping factor and phase parameters respectively.

2.2. Synthesis

The synthesis of the complete signal, X[m], using this model provides a good approximation to the original signal when the input consists of a small number of sinusoids. However, when the number of sinusoids is large, the time-domain minimization does not perform as well and leads to audible artifacts in the synthesized signal. Furthermore, the amount of data describing the parameters is manageable, but somewhat large due to over-representation. The solution to utilizing the benefits of the high-resolution transform in this context is to apply the well-known method of cubic interpolation for phase unwrapping from frame to frame [4]. As far as the aural quality is concerned, this method generates smooth partial trajectories and performs better than the time-domain minimization synthesis. In this case, the analysis window is hopped in a densely overlapped fashion that unfavorably adds to the computational cost.

Many frequency components are estimated as a result of the application of the matrix-pencil to each frame, but as would be readily expected not all are equally pertinent to the synthesis of the signal. The components of interest for the synthesis are those that have relatively small damping factors (< 0.007) and sufficient amplitudes (>-60dB relative to the peak amplitude). Hence, by pruning the very rapid transients and low amplitude components in each frame the most prominent components are obtained. Note that the damping factors can be negative during the attack portion of a signal.

The frame parameters are calculated for each frame by hopping the analysis window by approximately 10 msecs. After the relevant components have been calculated, partial tracking is performed. A slightly modified version of the traditional birth-death strategy [4][5] is employed. In this stage, additional filtering is required as some isolated spurious components might have escaped the steady state filtering explained above. This is done by ensuring that each frequency track is longer than a certain duration (typically 40 msec.). The tracking algorithm uses a combined amplitude-frequency metric to measure the distance between spectral components. This is a weighted metric that links the closest two components in the time-frequency-amplitude space. The partial tracking algorithm also performs a two-way distance verification to ensure that the two linked components are the closest ones in both forward and reverse directions in time.

3. EXAMPLES

Figure 1 shows the initial 1.3 seconds of the time waveform of a guitar playing the notes C3-D3 simultaneously, which roughly corresponds to 130.8 Hz and 146.8 Hz respectively. Note the overall beating pattern in the time waveform. Figure 2 shows the resynthesized signal using the described method. The identical beating pattern in the envelops can be observed by comparing the two figures.

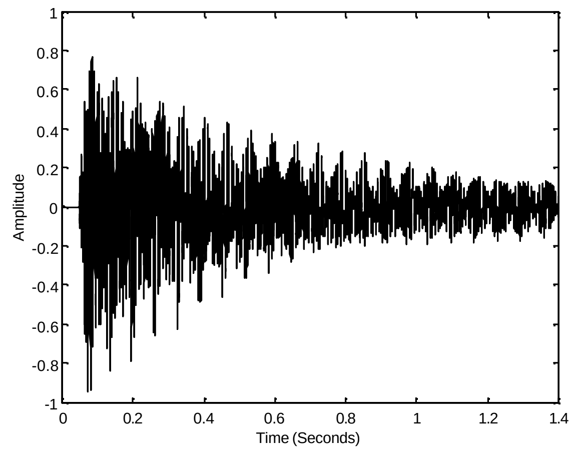

Figure 1. Input signal : Guitar C3 and D3 played simultaneously.

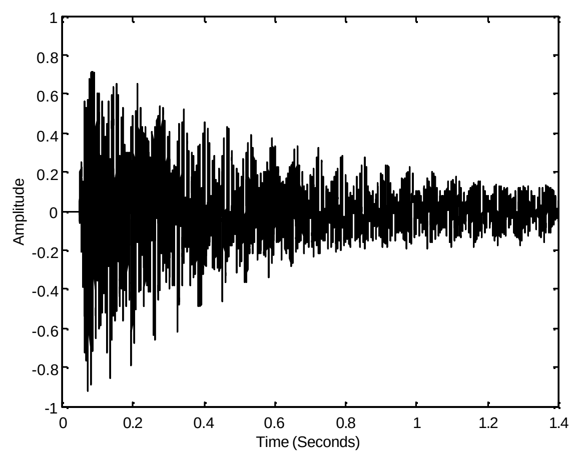

Figure 2. Time waveform of the resynthesized sound in Figure 1.

The same guitar signal is bandpass filtered to analyze the frequency region into which the ninth harmonic of C3 and the eighth harmonic of D3 fall. In this specific case, the partials are approximately 8 Hz apart. Figure 3 shows the time waveform of the resynthesized signal corresponding to the narrow frequency band surrounding these partials. The resynthesis is based on the time-domain minimization as described in equations 1 and 2 and the envelop seen in this figure is identical to the filtered input signal. The amplitude scale in this figure is one tenth that of figures 1 and 2.

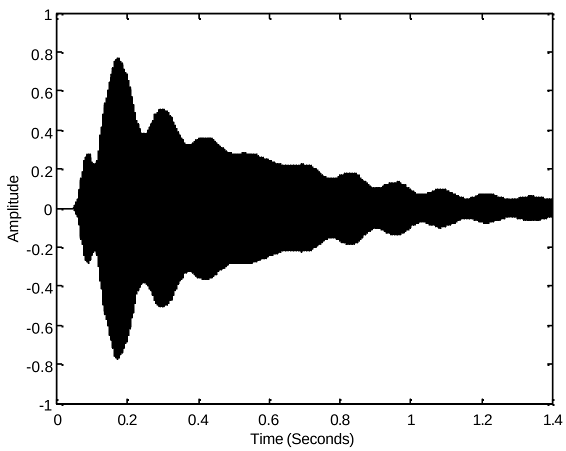

Figure 3. Resynthesized signal of ninth harmonic of C3 and the eighth harmonic of D3 of the guitar tone in Figure 1. Full scale in this figure corresponds to one tenth the full scale in figures 1 and 2.

4. DISCUSSION AND CONCLUSION

The major drawback of the matrix-pencil method is its computational cost. The method requires that the maximum number of sinusoids in the input signal be estimated. Execution time of the algorithm can be kept short for those signals known to have relatively few sinusoids by choosing the parameter values conservatively. On the other hand, for signals with many partials, the input has to be divided into subbands and the spectral analysis is performed on each subband later to recombined before going into the partial tracking stage. This is due to the limitation on the number of sinusoidal components that can be detected by the matrix-pencil analysis method, which is in turn bounded by the length of the analysis window.

The high-resolution front-end provides improved spectral resolution that opens possibilities for polyphonic sinusoidal analysis and modeling by resolving both isolated and close spectral components. The advantage of the high-resolution front-end is that it eliminates the harmonicity assumption (or equivalently the determination of the minimum frequency component spacing) on the input signal and can even cater to multi-voice inputs. The method has been effectively applied to musical signals with reasonably slow frequency variations to analyze and synthesize close and beating frequency pairs.

5. REFERENCES

[1] Izmirli, O., “A Hierarchical Constant Q Transform for Partial Tracking in Musical Signals” Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99), Trondheim, Norway, pp. 109-111, 1999.

[2] Beauchamp, J. W., “Analysis and Resynthesis of Percussive Sounds: Two Methods Compared” Proceedings of the International Computer Music Conference (ICMC99), Beijing, China. pp. 347-350, 1999.

[3] Laroche, J. “The Use of the Matrix Pencil Method for the Spectrum Analysis of Musical Signals” Journal of Acoustical Society of America, Vol. 94, 1958-1965, 1993.

[4] McAulay, R. J., T. F. Quatieri "Speech Analysis/Synthesis Based on a Sinusoidal Representation" IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 34, 744-754, 1986.

[5] Serra X.. "A System for Sound Analysis/Transformation/ Synthesis Based on a Deterministic Plus Stochastic Decomposition," Ph.D. Dissertation, Center for Computer Research in Music and Acoustics, Stanford, California., 1987.