Adaptive additive synthesis using spline based parameter trajectory models

Axel Röbel
IRCAM, Analysis-Synthesis Team, France
email: Axel.Roebel@ircam.fr

Abstract

We present the results of an analytical study concerned with the frequency resolution of our adaptive additive synthesis model. First, we derive the relation between the characteristics of the piece wise polynomial parameter trajectories of the model and the frequency resolution that can be obtained by means of adapting the model using a minimum error objective. Second, we present an analytical investigation of the problem to model signal resonances beyond the frequency resolution of the model. Based on the analytical description of the situation a new solution is proposed that leads to high quality additive models of non stationary sounds with dense resonances, i.e. choir or drum sounds, and provides increased robustness with respect to sound transformations.

1 Introduction

Additive synthesis is one of the major means for analysis/synthesis systems which, originally, has been restricted to only weakly non stationary sounds. Recent enhancements, however, successfully deal with non stationary and percussive sounds as well (Fitz, Haken, and Christensen 2000b).

Most of the known analysis algorithms for additive models use the short time Fourier transformation (STFT) to derive sampled parameters of the amplitude, frequency and phase of the model partials. Problems with this approach are the requirements to group the samples to obtain proper partial trajectories and to find a compromise between the usually inconsistent frequency and phase estimates. To prevent these problems a new adaptive approach has been developed (Röbel 1999). In contrast to the STFT based estimation techniques it adapts a piece wise polynomial trajectory model for amplitude and phase trajectories such that the model error is minimized. Due to the fact that the parameter trajectory model is explicitely and consistently used for finding the optimal model parameters no further heuristics are required and phase and frequency are a priori consistent.

In the following we will present two results of an analytical study of the model. The first is related to the frequency resolution that can be obtained by means of adapting a polynomial trajectory model. We will show that the frequency resolution is determined by the basic splines of the polynomial trajectory model. Second we investigate the case where the signals resonances to be modeled are dense and can not be resolved into independent partials due to insufficient frequency resolution. We derive a new formula that describes the relations between the amplitude and phase parameters of a set of non stationary cosines and the amplitude and phase of a single partial that exactly represents the set. Based on this new understanding of the problem we extend the adaptive algorithm such that the robustness of the synthesized sounds after parameter transformation is significantly increased even for difficult examples with dense resonances, i.e. a drum or a choir signal.

Note that, for distinguishing between the physical process that originates the signal and the mathematical model the terms resonances (signal) and partials (model) will be used in the following investigation.

2 The parameter trajectory model

The parameter trajectory model analyzed in the following is an improved version of the piece wise linear trajectory model used in the original version of the algorithm (Röbel 1999). To improve the convergence of the adaptive algorithm the frequency trajectory models of the algorithm have been replaced by phase trajectories. The advantage of the phase trajectories is related to the fact that non local operations have to be undertaken to achieve a local change in phase with a frequency based trajectory model. Additionally, the piece wise linear trajectory models have been replaced by piece wise polynomial trajectory models with arbitrary order.

The description of the polynomial trajectories by means of basic splines (B-splines) renders the mathematical treatment simple and straight forward (de Boor 1978) because a piece wise polynomial function $x(n)$ of order o can be expressed by linear superposition of B-splines of the same order following

$$x_o(n) = \sum_i B_i b_i(n). \quad (1)$$

Here B_i is the weighting parameter of the i-th B-spline of order o, $b_i(n)$. Note, that B-splines are functions with local support, hence they are non zero only in a connected and bounded region. Moreover, there exist different types of B-splines that obey increasing smoothness constraints on the boundaries of the supported region. For a B-spline that converges maximally smoothly to zero at the support boundaries the first $o-1$ derivatives are smooth with existing derivatives. To facilitate a less restrictive decay of a partial trajectories less smooth B-splines are used at the trajectory end points. A single partial is modeled as

$$P(n) = A(n)\cos(\Phi(n)), \tag{2}$$

where $A(n)$ and $\Phi(n)$ are the piece wise polynomial amplitude and phase trajectories. Using eq. (1) the partial model can be written as

$$P(n) = (\textstyle\sum_l (A_l b_l(n)))\cos(\sum_i \Phi_i b_i(n)). \tag{3}$$

Note, that if the same polynomial order is used for amplitude and phase trajectory the same B-Splines can be used, and, that a change of the polynomial order can simply be achieved by exchanging the B-splines b_i.

The resulting model $L(n)$ that represents the sound signal $S(n)$ is simply a sum of partials of the form described in eq. (3). The model parameters are adapted using a second order scaled conjugate gradient algorithm (Møller 1993) that minimizes the sum of squared error

$$E = \textstyle\sum_n (S(n) - L(n)). \tag{4}$$

3 Frequency resolution

For existing additive synthesis algorithms the properties of the analysis stage are characterized by means of the time/frequency resolution that can be obtained. This resolution is determined by the shape of the window that selects the analyzed block of samples, and its spectral main lobe width and side lobe height. The result described in the following section shows that the analysis properties of our adaptive algorithm can be expressed in terms of time/frequency resolution, too, and that this resolution is determined by the the basic splines b_i.

While the time resolution of a piece wise polynomial parameter trajectory is obviously related to the length of the polynomial segments, the determination of the frequency resolution requires further investigation. In the following we will estimate the impact of a small stationary perturbing cosine on the estimation of the partial parameters for a single non stationary resonance. Hence, the signal to be studied is

$$S(n) = A_s(n)\cos(\Phi_s(n)) + a\cos(wn). \tag{5}$$

The target partial parameters $A_s(n)$ and $\Phi_s(n)$ are assumed to match the polynomial trajectory model such the trajectories can be represented without error. Using this signal the influence of the disturbing cosine on the optimal partial parameters can be studied. Starting point are the equations that require the gradient of the error function with respect to the parameters to be zero at the global minimum. For $a = 0$ the global minimum is achieved for $A_s(n) = A(n)$ $\Phi_s(n) = \Phi(n)$. These zero conditions are now linearized with respect to a, and the model parameters A_i, Φ_i around the global optimum for $a = 0$. The linearized equations will approximately describe the change of A_i and Φ_i for small changes of a and can be used to determine the impact of the disturbing cosine on the partial parameters.

The mathematical investigation reveals, that the change of the optimal model parameter due to the disturbing signal, are linear combinations of all the products that can be build out of the disturbing signal, a single B-spline of the trajectory model and a harmonic function of the partials optimal phase trajectory

$$\begin{aligned} d_{i1} &= \textstyle\sum_n a\cos(wn)\cos(\Phi_s(N))b_i(n) \\ d_{i2} &= \textstyle\sum_n a\cos(wn)A_s(n)\sin(\Phi_s(N))b_i(n). \end{aligned} \tag{6}$$

These products can be interpreted in the frequency domain by noting that the multiplication between the B-spline, which has the form of a standard window function, and the disturbing signal results in a widening of the spectral peak related to the disturbing signal by means of a convolution with the B-splines Fourier spectrum. The following product and sum over n can be interpreted in the frequency domain as a sum over the product of the spectrum of the optimal partial and the convolved disturbing peak. Because the disturbing signal affects the model only after being windowed by means of a basic spline we conclude that the B-splines, that are used to represent the phase and amplitude trajectories, determine the frequency resolution of the adaptive algorithm in a similar manner as the window function determines the frequency resolution that can be obtained with STFT based algorithms. With respect to the adaptive model the following conclusions have to be drawn:

First, because the B-splines depend on the size of the polynomial segments they have to be considered as the link between frequency and time resolution of the algorithm. The B-splines for a piece wise polynomial of order o and segment length k can be derived by o-times convolving a rectangular window of width k with itself. Therefore, the frequency resolution that can be achieved using segment length k is comparable to the resolution obtained from a STFT based algorithm with rectangular window of size k.

Second, the cross talking of partials in different frequency bands is lowered with increasing polynomial order because the side lobes of the spectra of the B-splines are lowered with increasing order o. To achieve sufficient side lobe attenuation we generally select $o = 4$.

Third, to achieve optimal resolution for the adaptive process the data blocks used to calculate the model error and adapt the parameters should not cut a B-spline that is related to a parameter to be optimized. Otherwise the spectral representation of the effective B-spline will undergo a significant increase of the spectral main lobe width and the side lobe height which will reduce the frequency resolution of the analysis. Because the theoretical background has not been available yet, the previous implementation of the algorithm was not in accordance with the requirements and sub optimal frequency resolution has been achieved.

4 Beyond the frequency resolution

One of the problems for additive synthesis algorithms is the fact that the resonances that are contained in natural sounds are often to close in frequency to be resolved by the analysis algorithm. Different strategies have been developed to deal with the problem using some kind of noise processes that represents the dense resonances of the signal (Serra and Smith 1990; Fitz, Haken, and Christensen 2000a). Noise based models, however, perform badly if the number of resonances is not very high, as for example in choir signals.

Without any means to deal with the problem the original version of our adaptive algorithm performed badly if the frequency resolution was not sufficient to resolve all resonances. In these cases the sets of partials extracted from the sound were generally wildly frequency modulated and, while the sound quality of the resynthesized sound was high, the sound characteristics changed significantly whenever a transformation has been applied.

4.1 Modeling sets of resonances

The non stationary partial model used in eq. (3) supports an understanding of the model as a collection of amplitude and phase modulated partials, each of which will model all resonances that are close to their respective frequency trajectory. Therefore, an analytical formula was required that would describe the equivalence relations between a set of nonstationary resonances and a single modulated partial. While the relation is simple and well known if the set comprises only two stationary cosines, we are not aware of a mathematical equation describing the general case and matching the present situation. Therefore, the formula derived in the following provides new insights into the behavior of additive synthesis models applied to dense resonances. It may be used in the future to change a models internal representation from independent partials to modulated partials according to the requirements of an intended sound transformation.

Starting from a set of K resonances a single modulated partial is searched that exactly represents the set. The mathematical relation to be solved is

$$\sum_k A_k(n)\cos(\Phi_k(n)) = \hat{A}(n)\cos(\hat{\Phi}(n)). \tag{7}$$

Here the k-th resonance has amplitude and phase trajectories $A_k(n)$ and $\Phi_k(n)$ while the equivalent partials trajectories are $\hat{A}(n)$ and $\hat{\Phi}(n)$ respectively.

By simply replacing the left cosines in eq. (7) by means of the real part of an complex exponential, replacing the complex exponentials by their real and imaginary parts, summing independently over the real and imaginary parts of the partials and applying basic rules of complex arithmetic to re-extract the real part the desired relation is established. Starting from

$$\begin{aligned}\textstyle\sum_k A_k(n)cos(\Phi_k(n)) &= Re\{\textstyle\sum_k A_k(n)e^{j\Phi_k(n)}\} \\ &= Re\{R(n)+jI(n)\} = \hat{A}(n)\cos(\hat{\Phi}(n))\end{aligned} \tag{8}$$

one obtains easily amplitude and phase of the desired partial

$$\begin{aligned}\hat{A}(n) &= \sqrt{R(n)^2+I(n)^2}, \\ \hat{\Phi}(n) &= \text{sign}(I(n))\arccos(R(n)/\hat{A}(n)).\end{aligned} \tag{9}$$

To understand the erratic behavior of the partial model when the resonances are dense in frequency a very simple example has been studied. In this case the signal to be modeled contains three stationary resonances that are considered close in frequency with respect to the frequency resolution of the model. The values of the frequencies are $W_1 = 0.49rad$, $W_2 = 0.51rad$ and $W_3 = 0.505rad$. The amplitudes have been chosen to be $A_1 = 1.1$, $A_2 = 2$ and $A_3 = 0.8$. Using eq. (9) the equivalent partial has been calculated and its instantaneous amplitude $\hat{A}(n)$ and frequency $\hat{W}(n)$, which is the difference of the unwrapped phase trajectory with respect to time, are depicted in fig. 1. The figure reveals the fact that the instantaneous frequency is heavily modulated and moves out of the frequency domain of the resonances whenever the momentary amplitude is close to a local minimum. The deviation of the the momentary frequency in this case is quite large, more than 100% of the band width of the frequency range of the resonances. From the cases studied so far and from a simple argument related to the summation of the rotating complex pointers it can be concluded that in general:

For a maximum of the amplitude $\hat{A}(n)$ all pointers need to be aligned and in this case the frequency (speed of rotation) of the summation of pointers equals the sum of frequencies of the single pointers weighted by their relative length

$$\hat{W}(n) = \textstyle\sum_k(A_k(n)W_k(n))/\sum_k(A_k(n)), \tag{10}$$

where $W_k(n)$ are the instantaneous frequencies of the underlying resonances. While there is no rigorous prove this appears to hold approximately true for all local maxima of the amplitude $\hat{A}(n)$.

In contrast to this the instantaneous frequency shows a maximum deviation from the average frequency for all local minima of $\hat{A}(n)$ with the amount of frequency deviation being larger for smaller values of teh minimum amplitude. This is weekly supported by the argument, that the phase of the summation of pointers is more easily affected by small changes of a single pointer if the sum of the pointers is small.

The important conclusion to be drawn from this example is that a stable tracking of a group of resonances by means of a single modulated partial can not be achieved, because the instantaneous frequency of the modeling partial occasionally has to leave the frequency region defined by the resonances currently modeled and will, with high probability, start modeling another set that includes resonances from the neighborhood. The switching between the groups of resonances is

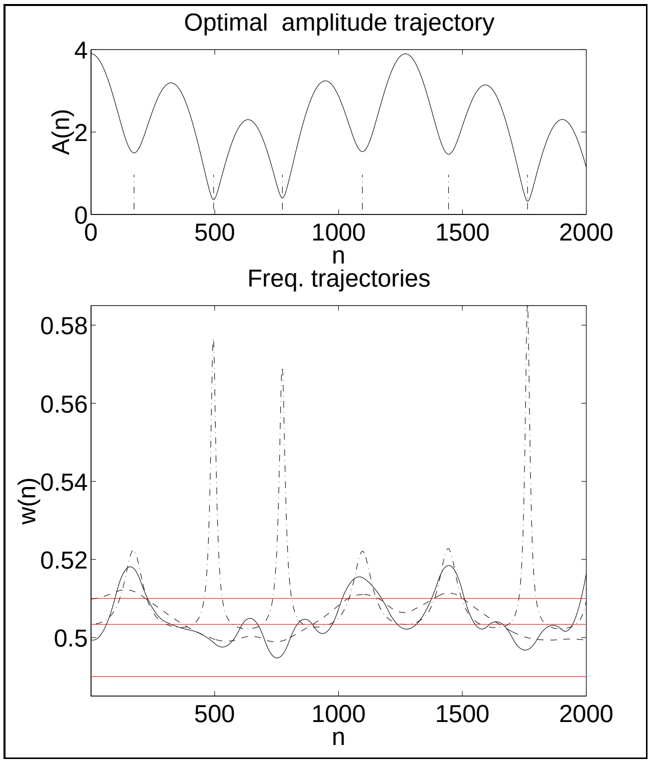

Figure 1: Top: Amplitude of the ideal partial model representing three stationary resonances. Bottom: Frequency of the ideal partial model (dash dots) and the adapted model with (dashed) and without regularization (solid). Segment size is 100 samples. Horizontal lines indicate the weighted average frequency (eq. (10)) and the frequency range of the resonances.

the reason for the strong frequency modulation that has been observed in all cases where the adaptive algorithm has been applied to signals with dense but fairly stationary sets of resonances. The resulting problem is twofold. First an interpretation of the partials found by the algorithm is hardly possible, and second, any manipulation of the partial trajectories will significantly effect the sound characteristics.

To solve the problem we rely on our conjecture that the instantaneous frequency will leave the frequency region of the resonances only while the instantaneous amplitude is small. Due to the small amplitude the peaks of the frequency modulation are expected to be subjectively not important and, therefore, the models phase trajectory can be constrained to follow only slow frequency changes. This can be achieved by adding a regularization term to the mean squared error minimization criterion that penalizes fast changes of the partials frequency. Regarding the frequency resolution of the model we consider fast a frequency slope that will change the frequency by more then the frequency resolution within the time of a polynomial segment. Note, that negative amplitude values are essential if the resonances can not be resolved because sign flips of the amplitude enable the phase trajectory to easily resynchronize whenever the frequency trajectory takes a short cut and, due to the constraints of a polynomial model, does not follow one of the fast frequency impulses.

The assumption that the frequency trajectory need not be modeled exactly has been verified by means of modeling some simple sets of resonances that are dense with respect to the models frequency resolution. The signal studied in fig. 1, for example, has been modeled with and without regularizing the frequency slope. The frequency trajectories derived in both cases are displayed in fig. 1. The original model can not track the ideal frequency trajectory due to the implicite constraint of the polynomial trajectory model,. Nevertheless, it is often far outside the modeled region. The regularized model has a smoother frequency trajectory which is always close to the frequency range of the resonances. Note, that the regularization should not be to strong, however, because it will prevent the model from following resonances with fast changes in frequency. Subjectively the sounds generated from both models are not distinguishable and very close to the original sound. For real sounds one can assume that the frequency region containing dense resonances is much larger than the bandwidth of a single modulated partial. Consequently, the model will engage a number of modulated partials to cover the region which will mask the small differences even further.

5 Results

The extended adaptive model has been applied to the set of sounds provided for the SDIF based additive synthesis comparison panel of ICMC 2000. The results obtained for the critical sounds support the theoretical results. In all cases high quality additive models with improved robustness with respect to sound transformations have been achieved. Examples for some of the most difficult sounds: drum, choir, harp will be presented at the conference.

References

de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer-Verlag.

Fitz, K., L. Haken, and P. Christensen (2000a). A new algorithm for bandwidth association in bandwidth-enhanced additive sound modeling. In *Proceedings of the International Computer Music Conference, ICMC'2000*, pp. pp.

Fitz, K., L. Haken, and P. Christensen (2000b). Transient preservation under transformation in an additive sound model. In *Proceedings of the International Computer Music Conference, ICMC'2000*, pp. pp.

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks 6*(4), 525–533.

Röbel, A. (1999). Adaptive additive synthesis of sound. In *Proc. Int. Computer Music Conference, (ICMC'99)*, pp. 256–259.

Serra, X. J. and J. O. Smith (1990). Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal 14*(4), 12–24.