Adaptive Additive Synthesis of Sound

Axel Röbel
Institute for Communication Sciences, Sekr. EN-8, Technical University of Berlin
Einsteinufer 17, 10587 Berlin, Germany
email: roebel@kgw.tu-berlin.de

Abstract

In the following article we investigate a new algorithm for additive sound synthzesis that extends the standard approach by means of explictely modeling a sound as a superposition of non stationary partials with time varying amplitude and frequency. Due to the extended model an adaptive parameter estimation is required. By means of a mathematical investigation we show, that it is possible to track a partials amplitude, frequency and phase by only adapting a subset of the model parameters and that a stable partial tracking requires the use of overlapping analysis blocks. We demonstrate the algorithm by means of additve synthesis of a synthetic and a piano tone.

1 Introduction

With the development of the phase vocoder techniques additive synthesis has been evolved into one of the most successful paradigms for resynthesis of natural sounds. Based on the short time Fourier transformation (STFT) the additive synthesis algorithms have the potential to resynthesize any kind of sound without error [1, 5]. However, a sensible transformation of the sound, for example a time stretch or a transposition, requires a sufficient match between the synthesis model and the physical characteristics of the sound source.

The problem has been realized in earlier investigations and is generally addressed by means of a linear interpolation of the parameters that are obtained from the STFT [4, 7]. This method has two problems. The first one is, that the linear interpolation of STFT parameters yields sensible results only if the sound parameters, especially the partial frequency, change sufficiently slowly. Due to the block averaging inherent to the STFT a change in frequency during the block will otherwise considerably affect the estimated amplitude. The second problem is due to the independent analysis of consecutive blocks which leads to the problem to match the respective partials across block boundaries.

To overcome these limitations we propose a new method for additive synthesis of sounds. From the above discussion of the current additive synthesis methods it becomes clear that the stationarity assumption that is a prerequisite for all Fourier analysis methods is the main reason for a failure of a successful application of additive synthesis for sound transformation. To be able to extend the additive model we, therefore, waive the Fourier transform and develop an adaptive algorithm that minimizes the error between model and sound signal. This allows us to use an extended non stationary additive synthesis model. With the proposed model we are able to extend all the partials from one block into the following such that a partial matching algorithm is not required.

In section 2 of this article we describe the extended additive synthesis model and our adaptation strategy and in section 3 we investigate the stability of the proposed partial tracking. In section 4 we discuss some algorithmic details of the new algorithm and in section 5 we shortly present a model for a piano sound that is able resynthesize the sound with no perceivable error and can be used to synthesize time stretched sounds with impressing naturalness. Section 6 concludes with a summary and an outlook on further work.

2 Additive Model

The additive synthesis model we are going to use in the following is of the form

$$\begin{aligned} x(n) &= \sum_{i=0}^{M} \hat{A}(n)\cos(\hat{\varphi}(n)) \qquad (1) \\ &= \sum_{i=0}^{M} (\hat{A}_i + \hat{a}_i n)\cos((\hat{W}_i + \hat{w}_i n)n + \hat{\phi}_i). \end{aligned}$$

All the $M+1$ partials have time varying amplitude $\hat{A}(n)$ with amplitude offset $\hat{A}_i$ and slope $\hat{a}_i$. The phase function $\hat{\varphi}(n)$ has as parameters the frequency offset $\hat{W}_i$ and slope $\hat{w}_i$ and the phase offset $\hat{\phi}_i$. The model explicitly assumes a non stationary evolution of the partials with a linear trend of amplitude and frequency. For brevity we do not denote the dependency of the partial parameter from the signal block number. To achieve proper transition of the partials from one block to the other we set the parameters $\hat{A}_i$, $\hat{W}_i$ and $\hat{\phi}_i$ at the beginning of a block to the respective values at the end of the previous block. Consequently, these parameters are only adapted for newly born partials.

To adapt the parameters we minimize the error between model $x(n)$ and sound signal $s(n)$ by means of gradient descend of the squared error over a block of samples

$$e = \sum_n d(n)^2 = \sum_n (s(n) - x(n))^2 \qquad (2)$$

The gradient of this objective function with respect to the model parameters is easy to calculate. For the slope parameters $\hat{a}_i$ and $\hat{w}_i$ for example, we find

$$\frac{\partial e}{\partial a_i} = \sum_n 2d(n)n\cos((\hat{W}_i + \hat{w}_i n)n + \hat{\phi}_i) \quad (3)$$

$$\frac{\partial e}{\partial w_i} = -\sum_n 2d(n)n^2 \sin((\hat{W}_i + \hat{w}_i n)n + \hat{\phi}_i). (4)$$

3 Adaptive Partial Tracking

Due to the restriction that amplitude, frequency and phase for a partial are determined by the respective partial of the previous block the proposed algorithm is recurrent and might become instable. To investigate into this problem we assume a sound signal that is representable by our linear model without error and try to determine the conditions that are required to decrease the error from the beginning of one block to the next. Note, that we distinguish between the true signals parameters and the model parameters by the hat that indicates an estimated parameter.

The analysis of the error equation eq. (2) is complicated and, therefore, we simplify the problem by means of studying the stability of amplitude and phase parameter tracking independently and start with the case of amplitude tracking. We assume that the signal to be modeled consists of a single partial with the true amplitude parameters A and a. We assume further that the frequency and phase parameters of the model are correct. These assumptions lead to the equation

$$e = \sum_{n=0}^{N-1} ((A - \hat{A}) + (a - \hat{a})n)^2 \cos(\varphi(n))^2. \quad (5)$$

For sufficiently high frequency the effect of the cosine is approximately a constant scaling of the error of the envelope and for the following investigation we simply neglect this constant. The amplitude offset of the model $\hat{A}$ is determined by the previous frame and, therefore, the only free parameter is the slope $\hat{a}$. Evaluation of the sum and finding the minimum with respect to $\hat{a}$ yields the result

$$\hat{a}_{opt} = a + 3\frac{A - \hat{A}}{2N+1} \quad (6)$$

With the optimal amplitude slope $\hat{a}_{opt}$ we find that the amplitude error at sample n

$$\Delta_A(n) = \Delta_A(0)(1 - \frac{3n}{2N+1}),\ \Delta_A(0) = A - \hat{A} \quad (7)$$

depends linearly on the initial error of the amplitude $\Delta_A(0)$. Moreover, for each n inside the block the amplitude error is smaller than the initial error with the minimum being zero at sample $\frac{2N+1}{3}$. We conclude that it is possible to track a linear amplitude over consecutive blocks by means of adapting $\hat{a}$ only, and that the fastest decay of an initial error is achieved if the consecutive blocks overlap by one third of the block size.

To analyze the stability of the phase tracking we assume that the amplitude slope is $a = 0$, that the signals amplitude is $A = 1$, and that the models amplitude parameters are already correct. In this case the error can be expressed as

$$e = \sum_{n=0}^{N-1} 2\sin(\frac{\Delta_\varphi(n)}{2})^2(1 - \cos(2\varphi(n) + \Delta_\varphi(n))), \quad (8)$$

where $\Delta_\varphi(n)$ denotes the error of the models phase function. For small $\Delta_\varphi(n)$ we get the approximation

$$e \approx \sum_{n=0}^{N} \frac{\Delta_\varphi(n)^2}{2}. \quad (9)$$

Again, the phase $\hat{\phi}$ and the frequency $\hat{W}$ are determined by the previous frame such that optimal slope of the frequency depends on the true frequency slope w and, moreover, on the initial phase error $\Delta_\phi = \phi - \hat{\phi}$ and the initial frequency error $\Delta_W = W - \hat{W}$ as follows

$$\hat{w}_{opt} = w + \frac{\Delta_\phi(10N+5) + \Delta_W \frac{15N(N+1)}{2}}{(2N+1)(3N^2+3N-1)}. \quad (10)$$

For large block size N and the optimal slope $\hat{w}_{opt}$ the phase and frequency error depend on the ratio between sample offset and block size $x = \frac{n}{N}$ as described by the following two dimensional linear equation

$$\begin{bmatrix} \Delta_W(x) \\ \Delta_\phi(x) \end{bmatrix} = \begin{bmatrix} (1 - \frac{5x}{2}) & -\frac{10x}{3N} \\ N(x - \frac{5x^2}{4}) & (1 - \frac{5x^2}{3}) \end{bmatrix} \begin{bmatrix} \Delta_W \\ \Delta_\phi \end{bmatrix}. \quad (11)$$

For a fixed block size N and sample offset x this equation can be interpreted as a linear system that describes the phase and frequency error at the beginning of the blocks for iterated application of the adaptive partial tracking algorithm with an overlap of consecutive analysis blocks of $N_o = N(1 - x)$. Further investigation of the stability of this linear system reveals the fact that the stability of the adaptive phase and frequency tracker depends solely on the size of the block overlap N_o. The magnitude of the eigenvalues of the partial tracking system are shown in fig. 1 together with the eigenvalues of the envelope tracking system eq. (7). Because the magnitude of the eigenvalue of the amplitude tracking system is always below 1 we conclude that amplitude tracking is always stable. Stable phase and frequency tracking, however, can be achieved only for sufficient overlap of the analysis blocks of at least $0.2N$. Due to the small magnitude of the eigenvalues for a block overlap of $0.3N$ we expect that this overlap will achieve the best tracking performance.

Due to the simplifications we have applied to be able to mathematically investigate our algorithm we can not expect the result to hold in all cases. One problem is the independent analysis of amplitude and frequency tracking behavior. Especially for fast changing amplitudes or frequencies the stability of the algorithm and the optimal block offset may change significantly from the results obtained. Another problem is the approximation of the sinus function in eq. (9). Here we expect that phase errors that are larger than π can change the stability of the algorithm. We have verified the stability of the algorithm with a simple test signal that consists of two

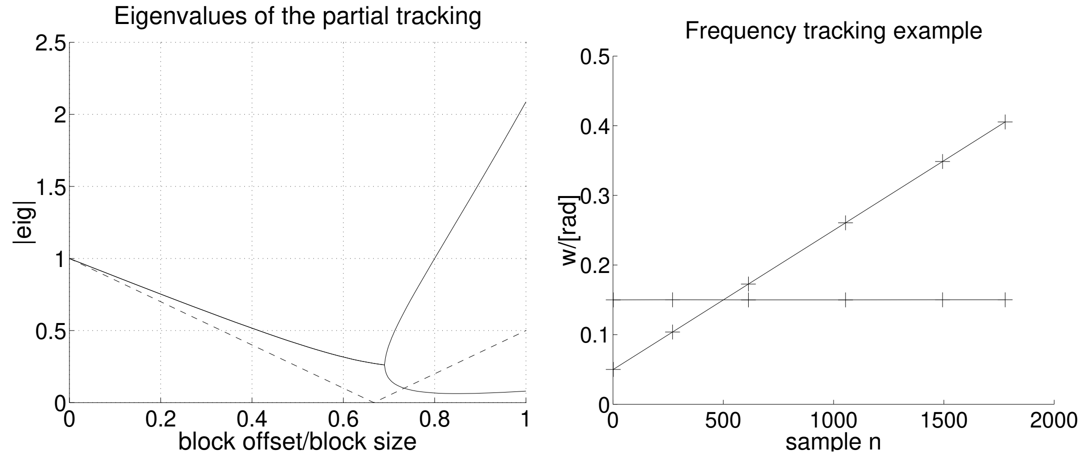

Figure 1: (left) The magnitude of the eigenvalues of the adaptive tracking algorithms depending on the block offset. While the amplitude tracking (dashed) is always stable the tracking of phase and frequency (solid) is stable for overlapping blocks, only. (right) Frequency tracking for a signal with two partials with crossing frequency (see text). Depicted is the frequency depending on the sample time for both partials of the model. A cross denotes the beginning of a block. The SNR is about 80dB and no significant deviation from the signals partials is detectable.

partials with linear frequency or amplitude trend

$$\begin{aligned} x(n) &= 0.5\cos(0.15n + 0.1) \qquad (12)\\ &+ (\ 0.3 + 3e^{-4}n)\sin((0.05 + 1e^{-4}n)n), \end{aligned}$$

such that the frequency of both partials cross after 500 samples. Note, that the second partial in contrast to the theoretical investigation combines a frequency and a amplitude trend. Due to the crossing frequency and due to the considerable slope of the frequency of the second partial a standard additive synthesis procedure that has been tested to model the above signal [6] fails to find a sensible model for the block with the frequency cross and achieves only bad resolution of the frequency and amplitude of the fast moving second partial. With the adaptive algorithm, a block size of 300 samples and an overlap of 100 samples, however, we are able to track both partials with no perceptual deviations from the true signal. In fig. 1 we present the instantaneous frequency of the model which have tracked the signal by means of solely adapting the parameters a and w of both partials. Note that the same experiment can be used to demonstrate the necessity of overlapping blocks, because for non overlapping blocks the frequency of both model partials become unstable.

4 Algorithmic considerations

To be able to apply the above adaptive strategy to the modeling of real sound signals a lot of further issues has to be considered. We need a method to initialize partials, to let the number of partials grow and shrink, to decide when to finish the adaptation for a block and to select a proper block size. At the present state we are not able to give a final guideline for all these issues, however, we are able to describe a fundamental algorithm that can be used as a starting point for further investigations and improvements.

For initialization of a new partials we use the standard parameter selection method of the additive synthesis algorithms. In detail we set the slope parameters to zero and determine the amplitude, frequency and phase from a local maximum of the Fourier transform of the current error. Because we expect that the maxima of the Fourier transform of the signal may not determine the number of partials required to model a non stationary signal we enlarge the adaptive model incrementally similar to the algorithms described in [2, 3]. The the maximum number of partials that may be used to model a block of samples is set as external parameter. After having selected this number of partials for modeling a block of samples we require that the partial that has been added in the last step has the smallest average amplitude of all partials. Otherwise, we delete the partial with the smallest average amplitude and select a further one. We constraint the gradient descent such that only positive frequencies between 0 and half the sampling frequency and positive amplitudes can be employed.

There may exist considerable energy in the neighborhood of a partial which is due to a non linear amplitude and/or frequency evolution of a signals partial. The solution to this problem is either a shorter block size, which increases the quality of a linear fit or an additional partial to model the lasting error. A possible way to implement the first solution is outlined in the next paragraph. The second solution has to be prevented because the training and tracking of two partials that have distance of less than $\frac{2\pi}{N}$ in frequency is not well defined. To select a new partial we, therefore, rate the correlation between the existing partials and the newly selected partial by means of a kind of psycho acoustically distance measure. For all the possible initial frequencies $\bar{W}_l$ with $l = \frac{k2\pi}{L}$ that are determined by the size of the Fourier transform L of the error signal we calculate the distance

$$c(\bar{W}_l) \quad = \quad \sum_{i=0}^{M} \frac{1}{N} \sum_{n=0}^{N} \qquad (13)$$

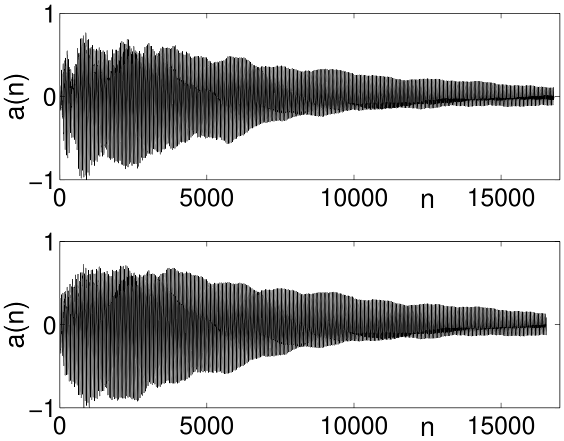

Figure 2: Comparison of the original (top) and resynthesized (bottom) piano sound signal.

$$\hat{A}(n)\cdot\ (1e^{-3} + \frac{1}{1+(N\frac{\hat{W}_i+\hat{w}_i 2n-\bar{W}_l}{2\pi})^2}).$$

and prior to selecting the maximum divide the error spectrum by $c(\bar{W})$. The devision yields a suppression of energy in the neighborhood of the existing partials, which due to psycho acoustic effects is less important for further modeling.

At the present time we did not implement a block size adaptation. However, we like to mention that the quality of the linear trend model can be validated by means of the gradient of the model parameters. For a minimum of the objective function eq. (2) the gradient over the block of samples is always zero. However, by means of investigation the local gradient at a sub block of samples we find that the sub block gradient will only be zero if the linear model is correct. Otherwise the sub block gradients are non zero and out weight each other such that the total gradient vanishes. Therefore, we intend to employ the behavior of the gradient over sub blocks to validate and control the block size of the adaptation.

5 Piano Sound Model

To give an impression of the possibilities of the new adaptive strategy we now present as an example the results for modeling a piano tone. The tone is a C_5 that has been recorded with 48kHz. The model is provided with maximally 14 partials and a block size of 500 samples. Due to the fast attack we decrease the block offset to $0.5N$. The original and the model signal are presented in fig. 2. By means of informal listening test we have verified that the differences between both signals are hardly perceivable even for professional listeners. It is interesting to note that time stretching the sound does not change the naturalness of the synthesized sound besides the fact that the time stretched stroke of the string sounds like a strucked and bended metal plate. Further investigation has to clarify, whether this is a physically plausible result or whether this is a kind of an artifact.

6 Outlook

We have presented a new algorithm for adaptive additive modeling of sound. A mathematical investigation of the partial tracking properties have revealed that only two out of the 5 signal parameters have to be adapted to achieve a tracking of partials and that stable tracking can be achieved with overlapping blocks only. We have discussed further details of our algorithm and present two examples for modeling a synthetic signal and a natural piano sound. Problems that have to be addressed is an improvement of the prevention of the modeling of a single signal partial by multiple model partials and an automatic adaptation of the block size.

References

[1] J. B. Allen and L. R. Rabiner. A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11):1558–1564, 1977.

[2] B. Edler, H. Purnhagen, and C. Ferekidis. ASAC - Analysis/synthesis audio codec for very low bitrates. Preprint 4179, *100th AES Convention*, 1996.

[3] E. B. George and M. J. T. Smith. An analysis-by-synthesis approach to sinusoidal modeling applied to the analysis and synthesis of musical tones. In *Proc. of the ICMC*, p. 356–359, 1991.

[4] R. J. McAulay and T. F. Quatieri. Speech analysis-synthesis based on a sinusoidal representation. *IEEE Trans. on ASSP*, 34(4):744–754, 1986.

[5] J. A. Moorer. The use of the phase vocoder in computer music applications. *Journal of the Audio Engineering Society*, 26(1/2):42–45, 1978.

[6] X. Serra. *Musical signal processing*, chapter Musical Sound Modeling with Sinusoids and Noise, pages 91–122. Studies on New Music Research. Swets & Zeitlinger B. V., 1997.

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