Adaptive additive modeling with continuous parameter trajectories

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

This preprint appears in: IEEE Transactions on Speech and Audio Processing, 14:4, pp. 1440-1453, 2006. Copyright IEEE.

***Abstract*— This article investigates into the estimation of time varying amplitude and phase trajectories of sinusoidal signal components. The new algorithm adaptively optimizes the parameters of a smoothly connected piecewise polynomial trajectory model. A mathematical analysis is presented that relates the user selected meta parameters of the trajectory model (polynomial order, segment size, and smoothness at the junctions) to the analysis properties of the adaptive algorithm. It reveals new insights into the relationships between the meta parameters and the resulting time/frequency resolution of the estimate. Moreover, it is shown that for efficient optimization the phase trajectory needs to be represented in a specific form. A new approach to address the bias/variance tradeoff of the polynomial phase trajectory model by means of regularization is presented and a complete adaptive analysis/synthesis system for sinusoidal sound components is proposed. The adaptive analysis system is investigated by means of simple tracking experiments to demonstrate the effect of the smoothness constraints and compare the results with a standard STFT base frequency estimation technique and known Cramer Rao bounds. The potential of the adaptive strategy for the modeling of sinusoidal transients is discussed and it is shown that it achieves similar transient quality as a previously proposed method, however, with considerably lower model error. Two examples for modeling real world signals are discussed.**

I. Introduction

The estimation of the parameters of sinusoidal components from an observed signal is a major step for many signal processing applications. One of the main applications in audio signal processing are additive analysis/synthesis systems. These are trying to represent a given sound signal, or at least part of it, by means of the superposition of time-varying sinusoids. Additive analysis/synthesis has been successfully applied to speech [1], [2], [3], [4] and music [5], [6], [7], [8]. The analysis of a sound signal in terms of a sinusoidal model brings up a number of issues related but not confined to these questions: what is a sinusoidal component, what part of the signal should be represented by sinusoids, what other signal models (for example noise) can be combined with the sinusoids to achieve an efficient representation with meaningful parameters, how are the parameters of the sinusoids represented and estimated, and how are the meta parameters of the analysis/synthesis system selected. The present article will be concerned mainly with the problem of representation and estimation of the sinusoidal parameters and with the selection of the meta parameters of the analysis/synthesis system.

Before the investigation is started it is necessary to define the notion of a sinusoidal component. In the following we consider a sinusoidal component as a nonstationary sinusoid of the form

$$P_k(n) = A_k(n)\cos(\Phi_k(n)), \tag{1}$$

where k is the identifying index of the sinusoid in the model, n is the discrete time and $A_k(n)$ and $\Phi_k(n)$ are the amplitude and phase trajectory of the sinusoid. The amplitude trajectory is constrained to be bandlimited. Even with this constraint the definition of a sinusoidal component is not sufficiently restrictive, because a simple Fourier transform of a time limited signal could be interpreted as a collection of constant frequency sinusoidal components. Likewise a wideband frequency modulation (FM) [9] of a single sinusoid using modulation frequency in the audible frequency range would perfectly match this definition of a single sinusoidal component. In both cases, however, the parameters of the sinusoids obtained are not related to the perceived sound characteristics, such that an intuitively meaningful control of the signal parameters is difficult. Therefore, the parameters of a sinusoidal component should obey a simple relation to the perceived sound characteristics.

The detection of sinusoidal components and the estimation of their parameters is usually based on the analysis of spectral peaks in the short time Fourier transformation (STFT) of the signal. As a result the values of amplitude, phase and frequency are available only at the frame centers of the STFT and the problem to connect and interpolate the parameter trajectories arises [5], [10], [1]. While the use of the STFT for parameter estimation is computationally effective the need to heuristically connect and interpolate the parameters is a significant drawback.

It has been stated very early [3] that estimation of the parameters of an additive model by means of minimizing a *meaningful cost function* would be a very promising approach. A first step towards an adaptive algorithm was the QUASAR signal model [8]. Similar to the model proposed in the following the QUASAR model starts with the specification of a trajectory model, which is a smoothly connected piecewise polynomial function. The optimization procedure, however, is based on an intermediate representation of phase and amplitude trajectories obtained by means of heterodyne filtering, and requires the sinusoids to stay close in frequency to an initially

selected nominal (or center) frequency. Because the nominal frequency of a sinusoidal component can never change this requirement is quite restrictive and the major benefit of the continuous phase trajectory model, which is the improved representation of nonstationary sinusoids, is not exploited. Because the analysis properties of the adaptive model and their relation to the polynomial order have been unknown, the size of the polynomial segments and the polynomial order of the QUASAR model had to be selected without taking the analysis properties into account.

A further adaptive approach that has been formulated in a Bayesian framework has been presented recently in [11]. In this case the model components are quasi-harmonic sets of sinusoids that are adapted to represent quasi-harmonic sound sources. The approach appears to be promising to solve the difficult problem of source separation. However, the parameter trajectory model that has been used is very limited and supports sinusoids with fixed frequency, only.

In the following article we will derive a clear understanding of the analysis properties of analysis/synthesis systems that are based on the minimization of the mean squared error (MSE) of sinusoidal functions with piecewise polynomial parameter trajectories. A detailed mathematical analysis of the global minimum is presented which proves the need for a specific representation of the phase trajectory of the model. Furthermore, the relations between the meta parameters of the trajectory model (the polynomial order and the segment size), and the resulting frequency and time resolution are established. Based on the new theoretical insights an iterative adaptive estimation procedure is proposed. The goal of this algorithm is the extraction of the parameter trajectories of the sinusoidal components of the signal such that it may be used as an replacement of the sinusoidal module of existing additive synthesis environments. Because the proposed method will represent only the sinusoidal components further modeling of the residual is required [7], [8]. The adaptive analysis significantly reduces the part of the sinusoidal energy that leaks into the residual such that the noise model may better match the residual.

Note, however, that due to the nonlinear adaptation involved the proposed algorithm is computationally much more demanding than the STFT based parameter estimation procedures. With the current implementation in MATLAB the computation required is in the order of 2000 times real-time such that an application is proposed only if minimization of the parameter estimation error is crucial

A new means to handle the bias/variance tradeoff for polynomial trajectory models [12] is proposed. It consists of extending the objective function by means of regularization terms that control the smoothness of the phase parameter trajectory. By means of simple tracking experiments we will compare the results obtained by means of adaptive parameter estimation with a traditional STFT based estimator and will discuss the relation to the known Cramer Rao bounds of the frequency estimation error.

The article is organized as follows. In section II we present the parameter trajectory model that is used and give a short introduction into the theory of B-splines. In section III we will explain, how the meta parameters of the piecewise polynomial trajectory model affect the time and frequency resolution of the estimation and in section IV we present a description of the optimization algorithm that is used and investigate into the amplitude scale dependency of the adaptive optimization. A complete algorithm for detection and adaptive optimization of sinusoidal components is presented in section V. In section VI an experimental investigation into the tracking performance of the adaptive algorithm is presented and the relation to the Cramer Rao bounds for frequency estimation is discussed. In section VII the problem of modeling attack transients is addressed and a simple yet effective strategy to reduce the pre-echo of the model is presented. Experimental results obtained with real world sounds are described in section VIII and an outlook on future developments concludes the article in section IX.

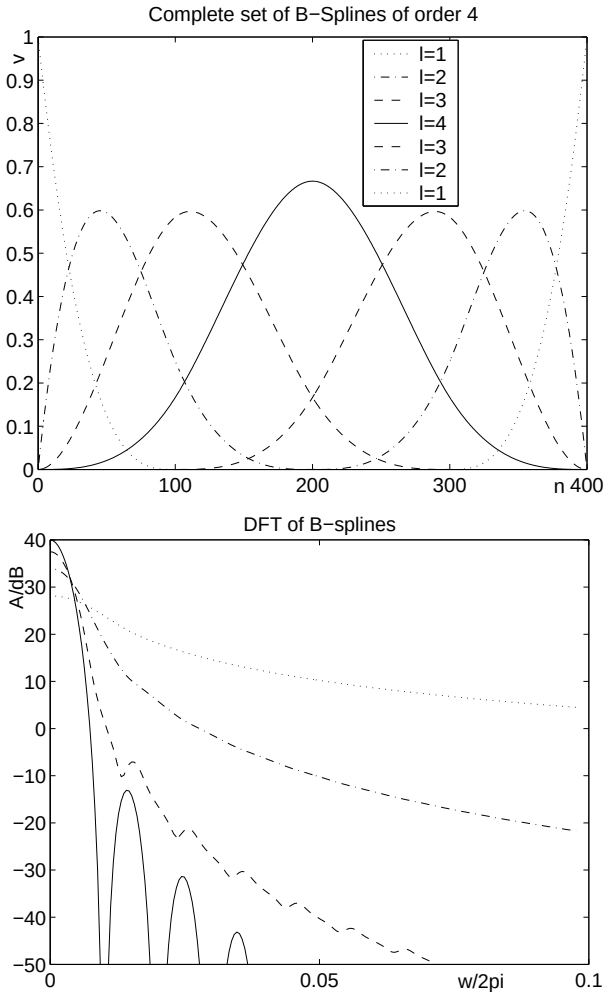

Fig. 1. A complete set of B-splines and their discrete Fourier spectra for a trajectory of length 400 samples. The B-splines shown have spline order $o = 4$ and segment length $M = 100$. At each side of the trajectory 3 zero size segments have been inserted to remove any smoothness constraints at the endpoints of the trajectory. Note that, while the B-splines cover always 4 segments some of the segments may have length 0 such that the related B-spline will be shorter. The smoothness of the B-splines is increasing with the number l of nonzero length segments that are covered. For longer trajectories only the number of maximally smooth B-splines ($l = 4$) would be increased.

II. The parameter trajectory model

In the following section we present the piecewise polynomial trajectory model that will be used for the amplitude and phase trajectories $A_k(n)$ and $\phi_k(n)$ in eq. (1).

Piecewise polynomial trajectories are commonly used in additive models to interpolate the trajectory parameters that have been estimated at the center positions of the STFT analysis frames. In that case, the polynomial order is determined by the available information and the size of the polynomial pieces, the segment size, has to be equal to the hop size of the analysis such that the polynomial pieces start and end at a frame center of the STFT [1]. For the adaptive model the parameter trajectory also consists of segments, the pieces of the piecewise polynomial function. In this case, the segments are fundamental and can be selected without referring to other parts of the model. The analysis frames of the STFT obtain their counterpart in the adaptive model in form of the basic splines (B-splines) that, as shown below, are implicitly defined by the user defined segments.

The representation of the piecewise polynomial trajectories by means of B-splines renders the mathematical treatment simple and straightforward [13]. A piecewise polynomial function $x(n)$ of order $(o-1)$ can be expressed by linear superposition of basic functions, the B-splines of order o, following

$$x(n) = \sum_i B_i b_i(n). \tag{2}$$

Here B_i is the weighting parameter of the i-th B-spline of order o, $b_i(n)$. The B-splines $b_i(n)$ are completely defined by their order o and the sizes of their segments. Due to the linear superposition it is obvious that the B-splines have to be piecewise polynomial of polynomial order $o-1$, too. Every B-spline covers o segments of the trajectory, some of which may have size zero. The B-splines are maximally smooth everywhere, besides at the locations where segments of size zero are covered. Each zero size segment inserted at a node position reduces the degree of smoothness at that position by one. To select a proper trajectory model we note that the order of smoothness of the parameter trajectories of a sound signal is generally unknown. There exist sound sources, e.g. vibrating bars and strings, flutes and pipes, that, besides during the attack, can be considered to have maximally smooth parameter trajectories. For others there may exist isolated points with reduced smoothness. As will become clear later, the reduction of the smoothness entails a reduction of the frequency resolution, and renders it time dependent. As a result, the parameter trajectories may become systematically modulated which makes parameter interpretation difficult. To circumvent these problems we will enforce the parameter trajectories to be everywhere maximally smooth, besides at the start and the end position. Consequently, the zero size segments will be present only at the start and the end of a trajectory.

An example for a complete set of B-splines that is necessary to represent a piecewise polynomial function with a possible step at the start and the end and smooth segment junctions for spline order $o = 4$ and segment size $M = 100$ is shown in fig. 1. The maximally smooth B-spline, covering o segments of size M will be used frequently and will be denoted as $BS_o(n)$. According to eq. (2) a single sinusoid with amplitude and phase trajectory represented by means of B-splines has the following form

$$P_k(n) = (\sum_l (A_{kl} b_l(n))) \cos(\sum_i \Phi_{ki} b_i(n)). \tag{3}$$

The free model parameters are the B-spline coefficients A_{kl} and Φ_{ki}. Note that in contrast to most of the known additive models the trajectory model used here does not restrict the amplitude to be nonnegative. This is suitable for the case when the amplitude trajectory to be modeled changes its sign. As an example one can imagine the case where two unresolved sinusoids have similar amplitude and need to be represented by means of a single amplitude modulated model sinusoid. The representation of the amplitude trajectory can be achieved either by means of a strictly positive amplitude and a phase step of size π or by a smooth phase and a smooth amplitude sign change. While smooth sign changes of the amplitude trajectory can be naturally expressed in eq. (3) by means of coefficients with varying sign, the representation of phase jumps or the restriction of the amplitude to be nonnegative would require complicated nonlinear extensions of the spline model and would significantly complicate the mathematical analysis and the adaptive optimization.

The additive model that is used to represent the sinusoidal components of a sound signal $S(n)$ is simply a sum of all sinusoidal components P_k and in a straightforward approach the model parameters could be adapted by minimizing the squared model error

$$E_0 = \sum_n E(n)^2 = \sum_n (S(n) - \sum_k P_k(n))^2. \tag{4}$$

III. Model parameter selection

For existing analysis/synthesis algorithms the properties of the analysis procedure are characterized by means of the time/ frequency resolution that can be obtained. The time resolution is determined by the size and shape of the analysis window, the frequency resolution by its spectral mainlobe width and sidelobe height. In this section the corresponding characterization of the properties of the minimum mean squared error solution of eq. (4) will be established. It will be explained how the time/frequency resolution of the representation of sinusoidal components using the piecewise polynomial trajectory model is determined by the B-splines.

Due to the time varying frequency of the sinusoids the mathematical analysis is involved. In the appendix we study the frequency resolution by means of deriving the impact of a perturbing signal on the minimum error parameters obtained for the representation of a single target sinusoid. It reveals, that the impact of the perturbing signal depends on the correlation between different parts of the target sinusoid and the perturbing signal. Because the parts of the target sinusoid to be correlated are determined by means of windowing it with the different model B-splines it can be deduced that the frequency resolution that is obtained for the minimum error solution of eq. (4) is determined by the Fourier transformation of the B-splines in quite the same manner as it is determined in the

standard analysis by the Fourier transform of the analysis window. It is well known that the time resolution of an algorithm is inversely related to its frequency resolution. Because the B-splines are the independent objects that are superposed to construct a trajectory, the size of the B-splines can be used as rough indication for the time resolution. This is confirmed by the results obtained in the appendix concerning the behavior of the model with respect to model errors. It is shown that these errors will be distributed over neighboring parameters with oscillating sign and decreasing amplitude. The trajectory error decreases to less than 10% within $\pm o$ neighboring segments. Note that for the modeling of sinusoidal transients a special post processing will be proposed to significantly reduce the pre-echo that is due to this distribution of the model errors. The results discussed so far allow us to draw some important conclusions for the use of the adaptive algorithm:

1) To derive a fundamental understanding of the spectrum of the B-splines we consider the maximally smooth B-spline $BS_o(n)$ for a constant segment length M. Using the recurrence relation [13]

$$\begin{aligned} BS_1(n) &= \begin{cases} 1 & \text{for} \quad n \in [0, M[\\ 0 & \text{else} \end{cases} \\ BS_o(n) &= BS_1(n) * BS_{o-1}(n)/M, \end{aligned}$$

we find that $BS_o(n)$ can be constructed by means of $(o-1)$-times convolving a rectangular window of width M with itself. Therefore, the Fourier spectrum of $BS_o(n)$ will be the power of order o of the spectrum of a rectangular window having mainlobe width of size $2\pi\text{rad}/(2M)$ and sidelobe attenuation $o \cdot 13dB$. Due to increasing sidelobe attenuation the impact of distant signal components on the parameters of a model sinusoid will be lowered with increasing spline order o. The price to pay is a decrease in time resolution due to the extended length $BS_o(n)$. To achieve sufficient sidelobe attenuation we generally select $o = 4$.

2) The B-splines that affect the frequency resolution during optimization are defined by their active part, which is the part that is used to calculate the model error. Consequently, parameters related to B-splines that are not fully covered by the analyzed signal segment should not be adapted because the effective B-spline is cut which significantly reduces the sidelobe attenuation.

3) It is well known that modeling phase trajectories with polynomial functions requires a bias/variance tradeoff. Increasing the polynomial order reduces bias because the model is less constrained but increases variance because the model may start to represent noise energy from the neighborhood of the sinusoid. The investigation in [12] reveals that phase and frequency estimation with completely unconstrained polynomial segments results in position dependent Cramer Rao bounds (CRB). For polynomials of order 4 the CRB varies by more than 12dB. This result is related to the fact that the B-splines that have to be used to create an unconstrained polynomial segment have significantly different frequency resolution. Due to the superposition of the B-splines, the frequency resolution will vary along the trajectory. As a result the estimated parameter trajectory becomes modulated due to the time varying impact of distant signal components. For the maximally smooth trajectory model that has been proposed here the B-spline $BS_o(n)$ can be used everywhere, besides at the start and end of the trajectory (see below). Therefore, the frequency resolution will be nearly constant and the systematic parameter variations are significantly reduced. The experimental investigation showed that the smoothness constraint reduces the impact of the polynomial order on the variance because the increased flexibility is accompanied by increased constraints. By means of adding regularization terms to the objective function eq. (4) further smoothness constraints can be created, that give rise to a continuous control over the variance of the phase estimation. For a spline order $o = 4$ possible constraints affect the second and third derivative of the phase trajectory[1]. Constraints of the first derivative of the phase trajectory are impractical because they restrict the range of possible frequencies of the sinusoidal components. For the second derivative, the frequency slope, we use a regularization term of the form

$$R_{2,k}(n) = (\frac{\partial^2 \Phi_k(n)}{\partial n^2} \frac{1}{F_2})^2 \quad \text{with} \quad F_2 = \frac{2\pi}{M^2}. \quad (5)$$

The regularization of the third derivative of the phase trajectory, the frequency curvature, is obtained by means of

$$R_{3,k}(n) = (\frac{\partial^3 \Phi_k(n)}{\partial n^3} \frac{1}{F_3})^2 \quad \text{with} \quad F_3 = \frac{2\pi}{M^3}. \quad (6)$$

The slope and curvature limits, F_2 and F_3, are related to the segment size M to ensure that the effect of the regularization term does not change with the size of the polynomial segments. The regularization factors are added to the objective function to be minimized which then becomes

$$E_R = E_0 + \sum_k \sum_n (\lambda_2 R_{2,k}(n) + \lambda_3 R_{3,k}(n)). \quad (7)$$

4) A problem of the initial version of the adaptive algorithm [14] were the physically unmotivated modulations within the border segments of the trajectories. As mentioned above these modulations are due to the fact that an inhomogeneous set of B-splines has to be used to represent trajectory borders. Some of these B-splines have significantly reduced frequency resolution and reduced sidelobe attenuation such that the parameter trajectories will be heavily affected by distant sinusoidal components. Remedy is simple for the amplitude trajectories because due to physical constraints the two least smooth B-splines are not needed to achieve high model quality. The phase trajectory, however, will generally be different from zero for both ends of a sinusoid such that all the B-splines displayed in fig. 1 are required.

[1]Throughout the article derivative with respect to n is understood to represent the derivative of the underlying time continuous function with respect to time at the position of sample n.

The oscillation of the phase trajectory at the trajectory borders is a serious problem. In order to track a sinusoid into a subsequent segment an initialization of the phase trajectory of the new segment is required that has to be derived from the phase trajectory of the previous frames. To improve the reliability of the initialization of the extension we rely on our analysis of the frequency resolution of the different B-splines. From fig. 1 we conclude that the results for the values at the first inner node of the trajectory, are only weakly influenced by the two border B-splines and provide a better basis to initialize the extended trajectory. Therefore, the extension of the phase trajectory is obtained by means of extending the trajectory with constant frequency slope starting from the last inner node position of the trajectory. While this procedure ensures reliable extension of sinusoidal parameter trajectories it cannot avoid the modulations that remain after adapting the initialized trajectory. To reduce those modulations the regularization terms mentioned above may be locally increased for the first and last polynomial segment.

IV. Parameter Optimization

To efficiently adapt the model parameters any second order optimization algorithm may be used. In this section we will briefly present the optimization algorithm that has been chosen and discuss a general problem of the optimization process that is related to amplitude scaling of the sinusoids.

The adaptive algorithm used for the following investigation is the second order scaled conjugate gradient algorithm proposed in [15]. It is a conjugate gradient algorithm [16] that has been modified to efficiently adapt nonlinear functions with many parameters. The basic idea is to avoid the line search in the conjugate gradient algorithm by means of estimating a local quadratic approximation of the objective function.

An important requirement for the successful application of an adaptive algorithm to optimize sinusoidal parameters is that the convergence properties do not change when the target signal is rescaled by means of a constant amplitude factor. Within audio signals there generally exist sinusoids with amplitudes covering three orders of magnitude or more and the behavior of the adaptive estimation should be similar for all of them. To highlight the problem we study a quadratic approximation of the global minimum of the objective function eq. (7) for the case that the target signal is a single sinusoid with parameter trajectories that can be modeled without error. Accordingly, the model contains only a single sinusoid ($k = 1$). The quadratic approximation of the objective function at the global optimum is completely described by its Hessian matrix H. For simplicity and without restriction of the generality of the results we are going to study a subset of two parameters A_{1i} and Φ_{1i}, only. We obtain

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} = \begin{pmatrix} \frac{\partial E_{err}}{\partial A_{1i}^2} & \frac{\partial E_{err}}{\partial A_{1i}\partial \Phi_{1i}} \\ \frac{\partial E_{err}}{\partial A_{1i}\partial \Phi_{1i}} & \frac{\partial E_{err}}{\partial P_{1i}^2} \end{pmatrix} \quad (8)$$

with

$$H_{11} = 2\sum_n b_i(n)^2 \cos(\Phi(n))^2 \approx \sum_n b_i(n)^2 \quad (9)$$

$$H_{12} = H_{21} \approx 0 \quad (10)$$

$$H_{22} \approx 2(\sum_n (A(n)^2 b_i(n)^2 \cos(\Phi(n))^2 + \sum_{I=\{2,3\}} \lambda_I (\frac{\partial^I b_i(n)}{F_I \partial n^I})^2) \quad (11)$$

$$\approx \sum_n A(n)^2 b_i(n)^2 + 2(\sum_{I=\{2,3\}} \lambda_I (\frac{\partial^I b_i(n)}{F_I \partial n^I})^2) \quad (12)$$

The approximations above are related to the fact that we have used the relations

$$\begin{aligned} \sum_n A(n)\cos(\Phi(n)) &\approx 0 \\ \sum_n A(n)\cos(\Phi(n))\sin(\Phi(n)) &\approx 0, \end{aligned} \quad (13)$$

that are due to the fact that the amplitude trajectories of the sinusoids are required to have limited bandwidth which is always smaller than the frequency of the sinusoid. Inspection of the equations H_{11} and H_{22} that determine the lengths of the two principal axis of the contour ellipsoid of the objective function reveals two problems. The first one is due to the fact that the relation between the regularization term and the error term of the phase derivative depends on the amplitude of the sinusoid. This is disadvantageous because the impact of the regularization would change after a simple amplitude scaling of the signal. The second one is due to the fact that the ratio of the principal axis is affected by the amplitude of the sinusoid. Note that multidimensional optimization of a (locally) quadratic objective function is performed most efficiently if all diagonal elements of the Hessian have similar magnitude. In this case the correct solution can be obtained in a single step. In the present case, however, the ratio of the magnitude of the diagonal elements of the Hessian matrix changes systematically with the amplitude of the target sinusoid. For large amplitude the gradient descend will consider adapting the phase parameters more important while for small amplitude adapting the amplitude parameters will be favored. To avoid these inconsistencies we modify the regularization terms by means of multiplying with the squared model amplitude trajectory

$$R'_{I,k}(n) = R_{I,k}(n) A_k(n)^2, \quad (14)$$

such that the impact of regularization will be independent of the amplitude of the sinusoid. Moreover, we use scaled B-splines $b'_i(n)$ given by

$$b'_i(n) = b_i(n) \frac{\sum_n b_i(n)}{\sqrt{A(n)^2 (b_i(n)^2 + 2(\sum_{I=\{2,3\}} \lambda_I (\frac{\partial^I b_i(n)}{F_I \partial n^I})^2))}}. \quad (15)$$

for constructing the phase trajectory in eq. (3). Scaling the phase B-splines ensures that the contour lines of the error function are independent of the target amplitude and always nearly circles. Because the scaling factors are not carrying any

meaningful information they should be considered fixed during adaptation. However, they are monitored, and whenever the optimal scaling factor differs from the current scaling factor by more than a factor two the scaling is changed and the conjugate gradient algorithm is restarted. The scaling of the phase B-splines is important to achieve good convergence for all sinusoids of the model, however, it does not affect the theoretical investigation, and therefore, we will ignore the scaling of the phase B-splines for the ongoing discussion.

V. The adaptive algorithm

The theoretic results presented so far have led us to considerably revise our first iterative adaptive additive model [14]. In this section we give an outline of the implementation of the algorithm that will then be studied with respect to its properties when estimating the model parameters.

The algorithm is iterative which means it adapts only a single sinusoid at a time. The reasons for this decision are:

- the difficulty to correctly handle the multiple solutions that exist when sinusoids close to each other are adapted together,
- the difficulty to correctly initialize the weak sinusoids from the signal spectrum without first removing the strong ones,

The phase trajectories use a full set of B-splines such that the sinusoidal phase function can start and end with arbitrary values. The amplitude trajectory is constrained to have 1st order smooth boundaries such that the two least smooth B-splines are not used. As objective function we minimize eq. (7) with the modified regularization terms given in eq. (14) and using scaled phase trajectory B-splines according to eq. (15). The signal segment that is used for training is a sliding window covering a maximum of K polynomial segments, where $K > o$ is a user selected parameter. For larger K the estimation results will be closer to the global optimum, however, at the expense of increased computational costs. As a reasonable compromise one may consider $K = 2o$. This will allow any parameter oscillations due to model insufficiencies to decay sufficiently before the parameters will be fixed. As discussed in section III only parameters related to B-splines that are fully covered by the current segment are adapted.

An overview over the basic steps of the algorithm is provided in the flowchart in fig. 2. The algorithm is implemented in a pseudo code listing in listing (I). As shown there the algorithm starts considering a signal segment of the size of $BS_o(n)$. This segment covers o polynomial segments. The main loop of the algorithm makes use of a STFT peak picking technique to detect and initialize new sinusoids using the standard analysis method described for example in [7]. To obtain consistently initialized sinusoids we use $BS_o(n)$ as analysis window. Because the adaptive algorithm will only produce reasonable results if the sinusoid is sufficiently covered we start by classifying the spectral peaks into transient and non-transient ones. As described in [17] this can be achieved by means of calculating the mean time [18] for each spectral peak. If the mean time of a single peak is above a threshold it is marked as transient. Because peak based processing is used only for partial initialization the transient detector considers only onset transients. The possible abrupt ending of sinusoids will be properly handled by the sinusoidal validation module (section V-A).

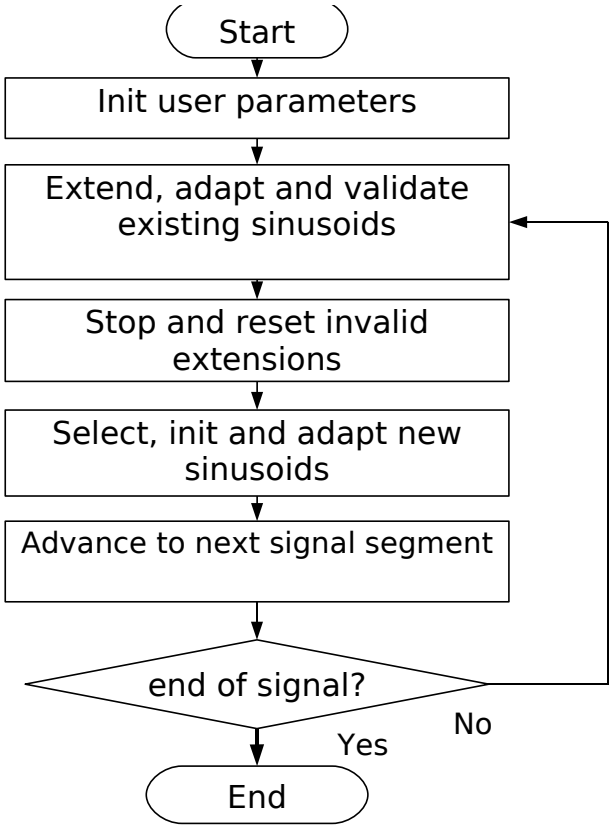

Fig. 2. Flowchart overview of adaptive algorithm.

Using the maximum peak of the spectrum we initialize a new sinusoid with fixed frequency and all amplitude coefficients besides the one related to the center B-spline to zero. The center amplitude coefficient is initialized according to the standard additive analysis procedure. The new sinusoid is adapted and after convergence is subtracted from the signal. It is checked for its validity as described in section V-A and marked according to the decision. Sinusoids that have been initialized from transient peaks are a priori invalid and are adapted with respect to amplitude coefficients only. Invalid sinusoids are subtracted from the signal to prevent an infinite loop when selecting the next sinusoid. These temporary components, however, will not become part of the model and are deleted after the adaptation of the current segment has been finished.

After subtraction the current error signal is calculated and used to select and initialize the next sinusoid. To ensure that the iterative algorithm will not use more than a single sinusoid for each spectral peak, the residual spectrum is masked by the existing sinusoids as described in section V-B. The iterative selection of new sinusoids repeats until the maximum number of active sinusoids requested by the user has been collected or a new sinusoid does not exceed a user supplied limit for its mean absolute amplitude.

After the current segment has been modeled the algorithm

```
K          = number of segments to adapt
signal     = signal to model
model      = empty
bslen      = length of smoothest B-spline
sigseg     = first bslen samples of signal
numparts   = max number of sinusoids in model
minamp     = minimum amplitude for a sinusoid to stay alive
dcmodel    = adapt sinusoid with fixed phase==0 to remove dc
signal     = signal - dcmodel
while sigseg not at end of signal
  modsort = sort model sinusoids according to mean amplitudes
  for all extsin in modsort
    extend amplitude trajectory by changing B-splines
      keeping old coefs and add new zero value coef.
    extend and adapt trajectory of extsin (see section III)
    determine validity of newsine (see section V-A)
    mark invalid sinusoids as stopped
  end
  cont := true
  while cont
    errsig   := sigseg - model
    ffterr   := FFT of last bslen samples in errsig
    mark transient peaks in ffterr
    determine masking thresholds in ffterr (see section V-B)
    select maximum non masked peak in ffterr
    newsine  = new and initialized sinusoid
    if new peak not transient
      adapt amplitude and phase of newsine until convergence
      determine validity of newsine (see section V-A)
    else
      adapt amplitude of newsine until convergence
      mark newsine as invalid
    endif
    if number of alive sinusoids < numparts
      add newsine into model
    else
      minsine = model sinusoid with minimum mean amplitude
      if ( mean_amp(newsine) > minamp
           and mean_amp(newsine) > mean_amp(minsine))
        mark minsine stopped, add newsine into model
      else
        cont = false
      endif
    endif
  end
  remove all newly born invalid sinusoids from model
  move sigseg to the next M samples
end
```

LISTING I

PSEUDO CODE DESCRIBING THE ORGANIZATION OF THE ALGORITHM.

extrapolates all valid sinusoidal components of the model into the next polynomial segment. The extended sinusoids are adapted in the order of their mean absolute amplitude using no more than the last K polynomial segments to adapt the parameters. All parameters of B-splines not covered by these last K segments are considered to have converged and will no longer be adapted. Extended sinusoids that do not match the validity criterion described below are marked as stopped and reset with the parameters they had prior to extension.

After the extension of all model components has been performed the selection loop restarts using the last o segments to select and initialize new sinusoids. Old sinusoids are stopped if the maximum allowed number of active sinusoids exists in the model and their average absolute amplitude over the last o segments is smaller then the corresponding value for a newly selected sinusoid. Due to the fact that newborn sinusoids are initialized using the last o polynomial segments they overlap the dead sinusoids of the last cycle by exactly $(o-1)$ polynomial segments.

A. Determining the validity of a sinusoidal component

Most additive analysis schemes include a mean to determine the end of a sinusoid. The correct detection of the end of a sinusoid is important because otherwise the parameter trajectory will be used to model unrelated sinusoids which may give rise to artifacts because the target parameter trajectories are inconsistent. In this section the criterion that is used to determine the validity of a sinusoid in the adaptive algorithm will be described.

The validity of a sinusoidal component is checked by comparing its amplitude trajectory to the amplitude trajectory obtained by heterodyne filtering the signal. The heterodyne filtering is done using the phase trajectory of the model sinusoid to be validated and applying $BS_o(n)$ as lowpass filter as follows:

$$a_c(n) = BS_o(n) * (s(n)\cos(\Phi(n))) \quad (16)$$

$$a_s(n) = BS_o(n) * (s(n)\sin(\Phi(n))). \quad (17)$$

Here $s(n)$ is the signal segment to model and $\Phi(n)$ is the phase trajectory of the model sinusoid. Note that the use of $BS_o(n)$ as a filter is motivated by the fact that the DFT of $BS_o(n)$ determines the impact of distant energy on the trajectory parameters. From $a_c(n)$ and $a_s(n)$ a time dependent complex phasor can be derived

$$a_{ref}(n) = \sqrt{(a_c(n)^2 + a_s(n)^2)}e^{j\,\mathrm{atan2}(a_s(n),a_c(n)))} \quad (18)$$

which is used as the reference amplitude trajectory. The reference trajectory has an amplitude and a phase component. If the model has managed to track a sinusoidal component of the signal, the reference amplitude should match the magnitude of the amplitude of the sinusoid and the phase should be either 0 or π according to the sign of the amplitude of the model sinusoid. The comparison is done by subtracting the two amplitude trajectories and calculating a running absolute normalized MSE according to

$$v(k) = \frac{\sum_{n=k}^{k+M-1}(a_{ref}(n) - A(n))^2}{\sum_{n=k}^{k+M-1} A(n)^2} \quad (19)$$

Whenever the maximum of $v(k)$ is larger than a threshold the sinusoid is considered to be invalid. The threshold has to be selected such that modulated sinusoids will not be cut. In all our experiments we have selected the threshold to be 3%.

B. Masking the residual spectrum

Due to the large energy difference between different parts of an audio signal it is common that the residual energy in a dominant audio band has larger amplitude than the sinusoids in weak audio bands. Due to masking effects, however, the residual energy will not be perceived. Because the proposed algorithm uses the DFT of the residual signal to initialize new sinusoids, proper masking is essential to prevent the addition of sinusoids that are irrelevant from a psycho-acoustical point of view. The basic idea proposed in [4] is to exclude all those

peaks in the residual from further modeling that are due to modeling errors of the sinusoids already present in the model. For the locally stationary sinusoidal model presented in [4] this masking can be simply achieved by means of excluding all those peaks in the residual that are closer to the model sinusoids than half the bandwidth of the mainlobe of the analysis window. In our case with time-varying amplitude and frequency trajectories of the model sinusoids a slightly more complex masking procedure is required.

In the following we denote the mainlobe of the spectrum of $BS_o(n)$ as $\Gamma(w)$. Because we are mainly interested to mask the residual close to each model sinusoid we use $\Gamma(w)$ as the basic masking function. For stationary model sinusoids with amplitude A_i and frequency w_i we can obtain a nearly equivalent masking effect as the one proposed in [4] by means of using $A_i\Gamma(w-w_i)$ as masking function related to the model sinusoid i.

To take into account the nonstationary amplitude and frequency evolution we need to redistribute the mask according to the effective impact of each instantaneous value of the amplitude and frequency trajectory. This impact will be controlled by the analysis window such that a sensible generalization computes a weighted average of the masking related to the instantaneous frequency and amplitude of sinusoid i using the analysis window as weighting function as follows

$$\Theta_i(w) = \frac{\sum_n |A_i(n)| BS_o(n) \Gamma(\frac{w-w_i(n)}{\beta})}{\sum_n BS_o(n)}. \qquad (20)$$

The masking threshold for frequency w and for the complete set of model sinusoids is simply the maximum value of the individual masking thresholds $\Theta_i(w)$. The scaling factor β allows to adapt the size of the masking range. For stationary sinusoids and with scaling factor $\beta = 0.75$, the masking threshold will be similar to the one proposed in [4]. For nonstationary sinusoids the masking will be lowered and spread over a larger frequency band. Note that for the following experiments the scaling factor is set to $\beta = 1$.

VI. Tracking sinusoids

In the following section we experimentally compare the frequency estimation error of the standard additive parameter estimator and the adaptive estimator described so far. The tasks consists of tracking single sinusoids in noise and the target frequency trajectories simulate important cases for real world sound signals. It will be demonstrated how the regularization parameters can adapt the trajectory model to specific trajectory characteristics.

For the following investigation we use analytic signals according to

$$x(n) = s(n) + r(n) = e^{i\phi(n)} + r(n). \qquad (21)$$

The signals comprise 40000 samples and the variance of the noise sequence $r(n)$ is adapted to achieve an SNR of 0dB within an analysis frame. The reference for frequency estimation is a standard additive approach using peak picking of the maximum in a 32768-point DFT and applying 3-rd order polynomial interpolation to obtain the frequency estimates [7]. The analysis window contains 2000 samples and the hop size is 500 samples. For the adaptive algorithm we use the segment size $M = 500$ and 4-th order splines as phase and amplitude trajectory model. The adaptive algorithm follows the description in section V using only a single model sinusoid and selecting $K = 15$. Note that the sinusoidal evaluation described in section section V-A has been switched off. To achieve a comparable setup we use $BS_o(n)$ as analysis window for the additive model. The standard method does not provide any reasonable estimates if the analysis window does not fully cover the signal, therefore, these estimates are not used. The CRB for frequency estimation of unconstrained polynomial phase signals have been derived according to [12]. The frequency error is specified in dB relative to the samplerate, and frequency values are specified as normalized frequencies such that the samplerate corresponds to frequency 1.

The first experiment is dealing with the case of a constant frequency sinusoid. This experiment simulates the frequency evolution that is common for example in plugged strings or vibrating bars. The frequency estimation errors obtained for the standard additive analysis and for the adaptive model with varying regularization parameters are displayed in fig. 3 (top). Due to the fact that an analysis window has been used, a common mean to reduce any bias from other sinusoidal components in the signal, the standard method is about 5dB above the CRB for constant frequency estimation. The CRB for the frequency error using a piecewise 3rd order polynomial phase function with segment size 500 and no smoothness constraints on the segment borders significantly depends on the position within the polynomial segment. The minimum is $26dB$ above the CRB for constant frequency estimation [12]. Due to the inherent smoothness constraints, the adaptive frequency trajectory achieves a frequency error which is well below the CRB of the unconstrained piecewise polynomial model even without regularization. For this case the error variance is about 14dB below the CRB of the unconstrained polynomial model and is only 7dB worse than the standard algorithm.

It is instructive to study how increasing the regularization according to eq. (7) will affect the results. The regularization reduces the degree of freedom of the model trajectory and, for the signal at hand, does not introduce any bias. As shown in fig. 3 for λ_2 and/or λ_3 being significantly larger than 10^{-2}, the smoothness constraints due to regularization start to dominate the inherent constraints of the spline model and, consequently, the estimation error decreases. For sufficiently large regularization the frequency estimation error drops well below the CRB of the standard analysis, which is possible because the adaptive method can make use of a larger part of the signal without introducing additional bias. For the situation at hand the lower limit of the error variance is given by the CRB of a constant frequency estimator that uses the data of the whole signal. For the given signal length the CRB is about -146dB. Due to the fact that the adaptive algorithm does not work globally but incrementally tracks the sinusoid never using a signal segment larger than 7500 samples to adapt the parameters this limit cannot be achieved. However, in the

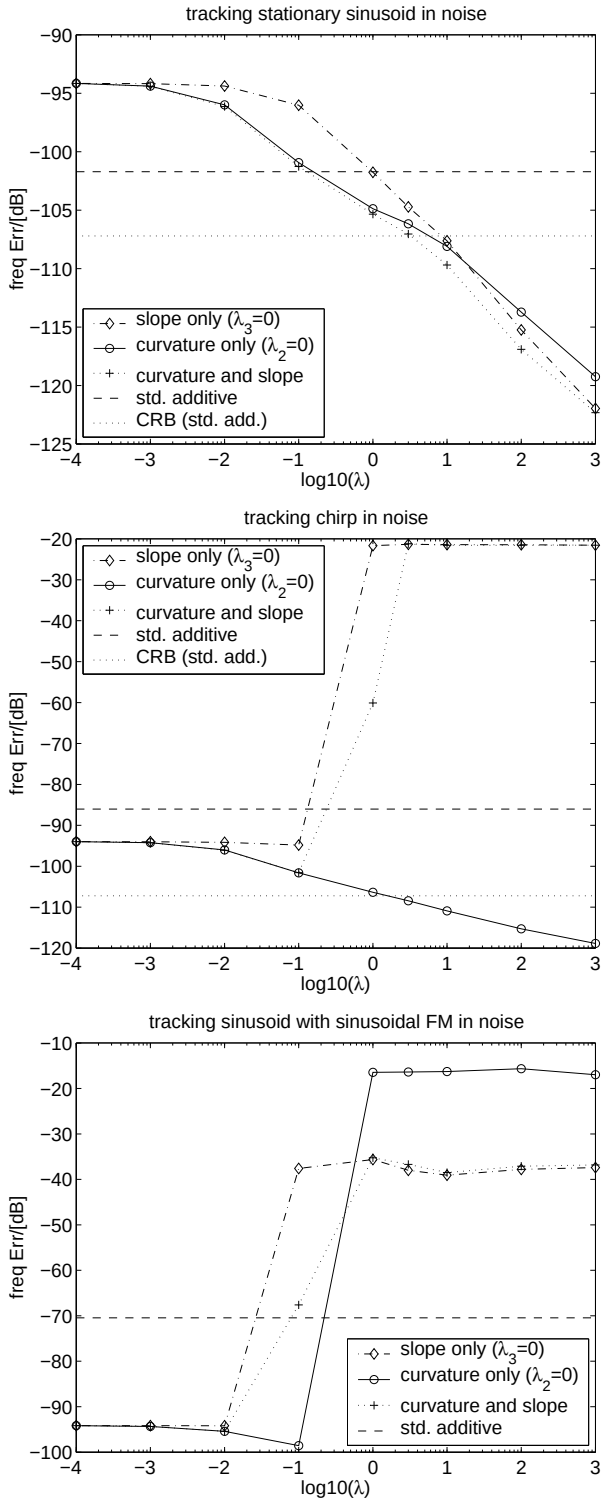

Fig. 3. Frequency error as a function of the regularization parameters for the cases: slope only ($\lambda_2 = \lambda, \lambda_3 = 0$), curvature only ($\lambda_3 = \lambda, \lambda_2 = 0$), and both ($\lambda_2 = \lambda_3 = \lambda$). Target is a sinusoid with constant (top), linear (center) and sinusoidal (bottom) frequency law in white Gaussian noise (SNR=0dB). The error obtained with a STFT based analysis algorithm and its Cramer Rao bound for frequency estimation are given for reference.

practical experiments the frequency estimation error comes rather close to this limit and saturates for $\lambda \approx 10^5$ at about -135dB.

In the center of fig. 3 the frequency error for tracking of a constant amplitude chirp signal is shown. The frequency slope is $\Delta_f = 4e^{-6}$ such that the frequency variation within the analysis window is not negligible. Accordingly, the normalized frequency changes from 0.16 to 0.32. The example has been selected to simulate sinusoids with considerable non periodic variation of the frequency trajectory. Note that the CRB for estimating chirp frequency trajectories with an unconstrained piecewise polynomial model is the same as for constant frequency signals. Due to the mismatch between the basis functions of the STFT and the chirp signal the frequency estimation error of the standard method is considerably increased. The adaptive model does not introduce any bias when modeling a chirp signal, and therefore the error variance is close to the previous case as long as no regularization is applied. Accordingly, for curvature regularization and increasing λ_3 we obtain approximately the same results as in the previous experiment. For slope regularization, however, increasing λ_2 above 10^{-2} results in increasing estimation error because the model is no longer capable to represent the target trajectory.

In the experiment shown at the bottom of fig. 3 the target sinusoid has a phase trajectory with sinusoidal frequency modulation. The center frequency of the sinusoid is 0.25 and the modulation frequency is $1.2421e^{-4}$. The extent of the sinusoidal modulation is a half tone, such that the sinusoid is resembling a sinusoidal component of a sound signal with vibrato. As shown in the figure, the frequency error of the standard additive model is further increased. If no regularization is applied the adaptive model still keeps approximately the same frequency estimation error that has been obtained in the previous experiments. We conclude that for the given parameters the B-spline polynomial model introduces only a negligible bias. Increasing the regularization, however, increases the bias such that first the models with regularized slope and at $\lambda_3 = 1e^{-1}$ the curvature regularized model can no longer track the signal. In this case the model degenerates into a nearly constant frequency sinusoid with mean frequency somewhere in the range of the frequency values of the target sinusoid.

From the results we conclude that the regularized adaptive frequency estimation with piecewise polynomial phase trajectories allows to adjust the properties of the trajectory model such that low variance of the frequency estimation can be achieved without restricting the polynomial order. The examples demonstrate that a wide range of common frequency laws can be handled by means of adapting the smoothness constraints for different orders of derivatives. The inherent smoothness constraints of the polynomial model did not introduce a significant bias in any of the cases that have been studied. For the sinusoidal FM the polynomial order of the model should probably be increased to be able to establish higher order smoothness constraints which would allow us to reduce the variance with less bias.

VII. Modeling sinusoidal attack transients

The representation of attack transients of resolved sinusoids is straightforward if rapid changes of the sinusoidal parameters are allowed. The analysis of fast changing parameters,

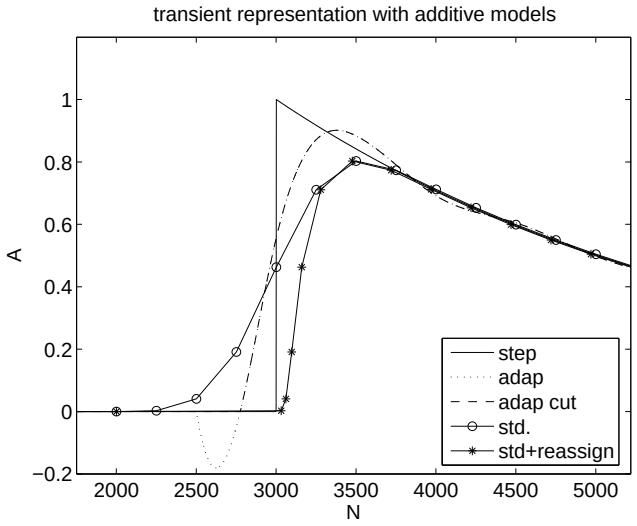

Fig. 4. Amplitude parameter trajectories obtained with different additive models (standard, reassigned, adaptive model with and without removal of pre-oscillations) for modeling a sinusoid with a step function as amplitude trajectory. To facilitate comparison only the amplitude trajectories are shown.

however, poses special problems for the parameter estimation algorithm because the high time resolution that is necessary to follow quickly changing parameters will compromise the frequency resolution. Due to the inherent conflict between time and frequency resolution it is common practice to use noise components to recreate the perceived attack [7]. The characteristics of spectrally shaped noise, however, are only suitable if the number of sinusoids that form the attack is rather large, which is not the case for example for string or bell sounds. Up to now, only few alternative approaches have been put forward to improve the representation of transients in an additive model. In the following some of the recent proposals will be discussed and will be compared with the representation obtained with the adaptive model. A simple trick is suggested that considerably improves the transient representation of the proposed model.

A recent approach to improve transient representation consists of extending the standard sinusoids plus noise model by means of a component that is especially dedicated to represent transients. In the sinusoidal model proposed in [19], [20] transients are directly represented by means of their spectra. The advantage of the spectral representation of transients is the high quality that will be obtained for simple re-synthesis and the fact that the model can be applied to represent noise transients as well. Because the transient representation does not provide sinusoidal parameters, however, it is difficult to compare it with the adaptive method such that it will not be discussed further.

The second approach to improve transient representation in a sinusoidal model is based on the relocation of the additive parameters using the reassignment operator [21]. The reassignment operator has originally been developed to increase the readability of the signal spectrogram [22]. It uses the phase spectrum to estimate the time frequency location of the signal component that is present at the time frequency location of the STFT. Time reassignment is closely related to the estimation of the mean time, the center of gravity of the signal energy, of the signal related to the current spectral peak [18]. For stationary signal components the time reassignment operator will be 0 such that sinusoidal parameters will be assigned to the frame center. For attack transients the emerging peaks in the STFT will initially be reassigned to the far right end of the analysis window and the reassignment offset will decrease with the window moving over the attack. In the early stage of an attack the analysis window does hardly cover the sinusoid such that the estimated parameters will suffer from reduced frequency resolution and can be simply discarded. Because later frames will provide better estimates for the same attack no information is actually lost [21]. Using parameter reassignment for the representation of attack transients of sinusoidal signals significantly increases the maximum slope of the amplitude trajectory which results in a perceptually very convincing attack representation. The drawback is, however, that the center of gravity of the signal energy has a systematic offset compared to the real amplitude trajectory, and therefore, the price that has to be payed for increased perceptual quality is a systematical increase of the residual energy of the time reassigned model.

A comparison of the amplitude trajectories obtained with the standard additive model with and without reassignment and the adaptive model is shown in fig. 4. The target sinusoid has an attack transient of the form of a simple step function with exponential decay. The model and analysis parameters are the same that have been used in the last section. The amplitude trajectory obtained with the standard additive model is maximally smooth with insufficient slope and starts nearly half a window before the step function. The amplitude trajectory of the reassigned model has increased slope but lies completely within the target amplitude trajectory such that the residual compared to the standard model is increased by 2.5dB. The amplitude trajectory of the adaptive model achieves a slope that is similar to the one obtained with reassignment, however, at the same time reduces the residual energy. The reduction of the residual energy depends on the position of the transient relative to the segment boundaries and for the current example ranges from -1.54dB to -3.65dB compared to the standard additive model. The figure shows an average situation achieving an reduction of the residual error by 2.3dB. The figure also demonstrates the effect of the transient detection which delays the initialization of the new sinusoid until it is sufficiently covering the newly initialized model sinusoid. Note the initial negative oscillation of the amplitude trajectory of the adaptive model, which prepares the model such that it eventually achieves a high slope. If the adaptive model would be restricted to positive amplitude values it would produce a trajectory similar to that of the standard model.

Making further use of the knowledge that the sinusoid just started a simple trick can be applied to partly suppress the pre-oscillations after the parameters have been optimized. Starting with the amplitude trajectory that results from the optimization we construct at most $(o-1)$ further amplitude trajectories by means of removing initial parts that extend up to the first, second and third zero crossing. From this set of at most o

different amplitude trajectories (including the original one) we select the one that achieves minimum error for the trained segment. As is shown in fig. 4 the resulting trajectory has its pre-oscillation removed without affecting the slope of the attack transient. The total reduction of the residual energy still varies with the transient position and now ranges from -1.74dB to -4.0dB. Due to the fact that this treatment is only applied for the initial part of a sinusoid and that it will only remove parts of the sinusoid that increase the model error there is no risk that parts of a real sinusoidal component may be removed.

VIII. Modeling real world signals

The algorithm described so far has been applied to many real world sound data files and has proven to be a favorable choice for signal representation when small model error is the main objective. As an example for a practical application we mention the development of a virtual replacement of a real acoustic pipe organ by means of an additive sampler [23] that has been carried out at our institute. To demonstrate that the algorithm works reliably with real world sound signals we will discuss the results obtained for two sounds from a database that had been collected to compare additive signal models using the Sound Description Interchange Format (SDIF) [24].

The two sounds have been selected to represent the two main problems for additive modeling that have been discussed in the current article: transients and nonstationary frequency trajectories. The transient sound example is a low pitch piano note with fundamental frequency of 65Hz. It can be found at [25] under the name *piano.aiff*. The sinusoidal part of the signal is difficult to represent with additive models due to the fact, that a long window is needed to resolve the sinusoids. However, to represent the attack the window should be short. The example concerned with tracking of frequency evolution is a singing voice signal that contains considerable pitch changes. This sound is accessible from the above mentioned database under the name *shafqat-derbari.aiff*

The initial segment of the piano signal and the residual signals of the sinusoidal model using an inharmonic standard STFT based parameter estimator and the adaptive method are depicted in fig. 5. For both additive models we allow the same maximum number of active sinusoids per time instant and have optimized the meta parameters such that they achieve minimum error. For the STFT based analysis procedure using a Hanning window the optimal window size is 0.041s while for the adaptive model with polynomial order $o = 4$ a segment length of $M = 0.0147$s has been selected. Based on the findings described in the present article this value could have been selected a priori because for all the values of M that have been tried the optimum one has the first zero of the B-spline spectrum located closest to the fundamental frequency such that the bias introduced by the neighboring sinusoids will be minimal. Knowing that the sinusoids have nearly constant frequency the regularization has been set to $\lambda_3 = \lambda_2 = 0.25$. The optimal window length of the standard model is shorter, however, in spectral domain the resulting mainlobe of the Hanning window and the B-spline are quite similar. The

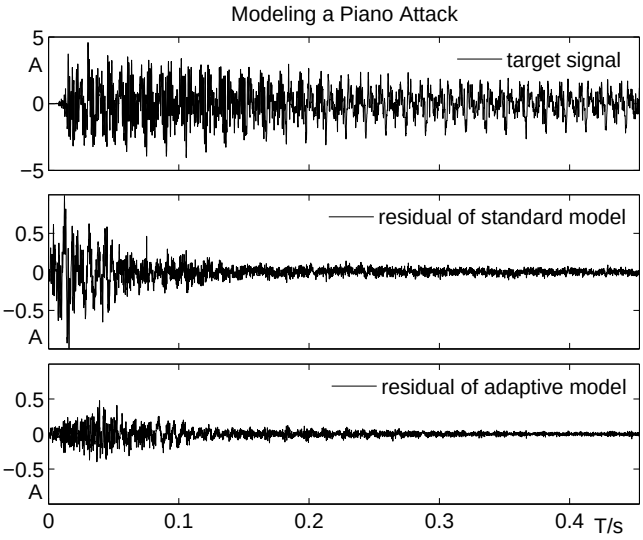

Fig. 5. Comparing the error of a standard additive model with the adaptive model when modeling a transient of a piano signal. Note that scales are different by a factor 5 for the original and the error signals.

residual energy of the complete segment shown is -20dB below the original signal for the standard additive algorithm (center) and -25.6dB for the adaptive model (bottom). The reduction of the residual error that is achieved by the adaptive model for the transient part in the first 100ms is 6.4dB while in the remaining part it still achieves 2.6dB reduction. The reduction in the stationary part is due to the fact that the Hanning window used for the STFT based model has less sidelobe rejection and that all sinusoidal parameters are estimated in a single analysis step which increases the bias for the multi-component signal. Note that, despite the increased precision, the re-synthesized signals are perceptually hardly distinguishable. The residual signal, however, is audibly easily distinguishable. First, because the attack part of the residual of the standard estimator has a slight tonal quality and second, because its stationary part contains significant beating due the fact that some low amplitude sinusoids have not been sufficiently resolved.

The second example, shown in fig. 6, demonstrates the tracking of time varying frequency trajectories in a singing voice signal. The sinusoidal frequency trajectories that have been found are laid over the spectrogram. In the spectrogram darker gray represents lower amplitude and the line thickness roughly represents amplitude. The long sinusoidal trajectories that represent stable sinusoids in the source signal are easily distinguished from the short ones that are due to noise. To improve the tracking of the nonstationary frequency trajectories the regularization has been slightly reduced to $\lambda_2 = 0.07, \lambda_3 = 0.2$. At the bottom of fig. 6 the original signal and the residuals of an harmonic additive model with STFT based analysis and the adaptive model are presented. The analysis window that provides best results with the harmonic model is a Blackman window of length 16ms. For the adaptive model best results have been obtained with a segment size of 5.7ms. For both models a maximum of 80 sinusoids is allowed

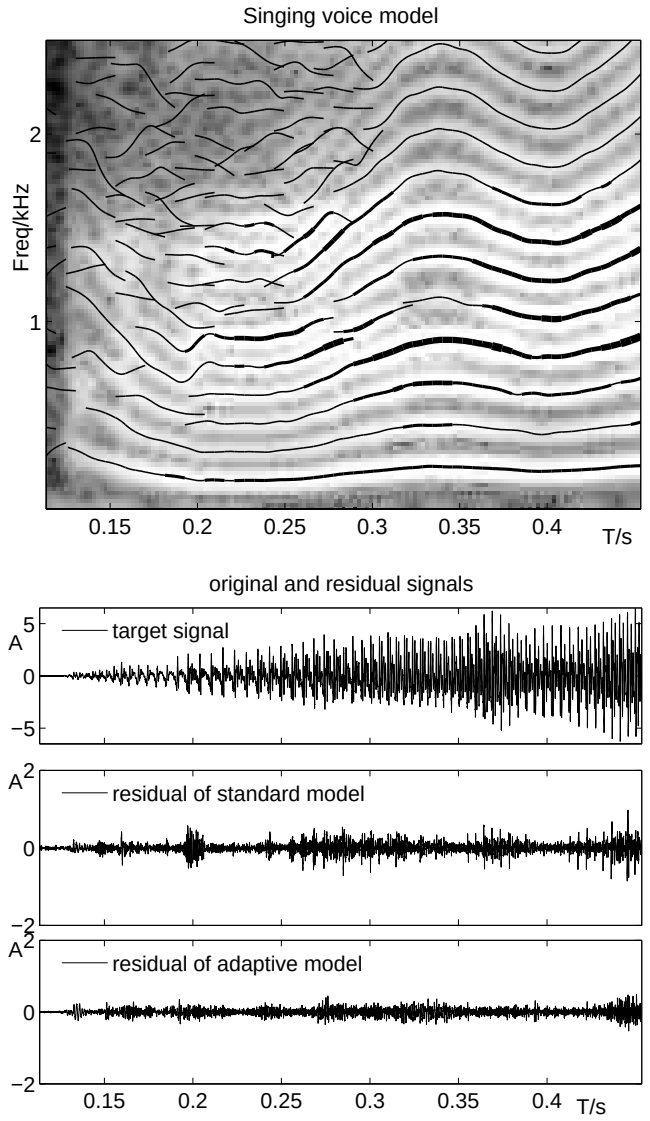

Fig. 6. Frequency trajectories of the additive signal model for the singing voice sound (top). The residual of the adaptive model is significantly smaller than the residual of the standard additive model.

at each time instant. The residual of the adaptive method is on average 2.1dB below the residual of the harmonic model. The comparison of the residual signals with the spectrogram reveals that the advantages are mostly related to improved tracking of sinusoids with fast changing frequency. An auditive comparison of the results shows that again the re-synthesized signals can not be distinguished. The tonality of the residual that is obtained with the adaptive model, however, is clearly reduced.

IX. Conclusions and Future work

A detailed investigation into adaptive estimation of sinusoidal parameters has been presented. The new insights into the analysis properties, notably the time and frequency resolution, of the adaptive optimization of a piecewise polynomial parameter trajectory model that have been derived, allow a straightforward selection of reasonable meta parameters of the model. The optimization procedure has been investigated and it has been shown that to achieve high efficiency the representation selected for the phase trajectory should be chosen as a function of the amplitude of the sinusoid under adaptation. We have proposed a new approach to handle the bias/variance tradeoff of piecewise polynomial phase trajectory models by means of regularization and have shown that the proposed regularization scheme allows to tune the model characteristics such that a variety of real world situations can be handled. It has been demonstrated that compared to STFT based parameter estimation the adaptive model achieves considerably improved representation of the transient part of sinusoids and considerably lower frequency estimation errors when tracking nonstationary sinusoids in noise.

Due to space constraints the current article has been limited to deal with resolved sinusoids. As has been show in [26] the regularized adaptive model has favorable properties for modeling limited numbers of unresolved sinusoids, too. The investigation into noise components, however, requires further discussion and will be the subject of an forthcoming article.

An interesting extension of the method would be the adaptive estimation of the parameters of quasi harmonic sets of sinusoids. If the complete, quasiharmonic set of sinusoids of a single instrument would be adapted simultaneously, advanced and physically motivated regularization of the different parameter trajectories with respect to their deviation from the harmonic model could be established. This could lead to significant improvements for the tracking of high order partials of harmonic sounds. Moreover, in combination with recently improved algorithms for the estimation of fundamental frequencies from polyphonic signals, this research direction will establish a new approach to separation of quasiharmonic sources from polyphonic signals similar to the Bayesian approach proposed in [11].

X. Acknowledgments

Part of the mathematical investigation of the adaptive algorithm has been performed during a research scholarship at the Center of Computer Research in Music and Acoustics (CCRMA) of Stanford University. The author would like to thank the people at CCRMA for their support, the Deutsche Forschungsgemeinschaft (DFG) for funding this scholarship and the reviewers for their valuable suggestions that substantially improved the original version of the article.

References

[1] R. J. McAulay and T. F. Quatieri, “Speech analysis-synthesis based on a sinusoidal representation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[2] T. F. Quatieri and R. J. McAulay, “Shape invariant time-scale and pitch modification of speech,” *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 497–510, 1992.

[3] J. S. Marques and L. B. Almeida, “A background for sinusoid based representation of voiced speech,” in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, 1986, pp. 1233–1236 (Vol. II).

[4] E. B. George and M. J. T. Smith, “Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model,” *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 389–406, 1997.

[5] J. O. Smith and X. Serra, “PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation,” in *Proc. Int. Computer Music Conference (ICMC)*, 1987, pp. 290–297.

[6] 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*, vol. 14, no. 4, pp. 12–24, 1990.

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

[8] Y. Ding and Q. Qian, "Processing of musical tones using a combined quadratic polynomial-phase sinusoid and residual (quasar) signal model," *Journal of the Audio Engineering Society*, vol. 45, no. 7/8, pp. 571–585, 1997.

[9] J. M. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *Journal Audio Eng. Soc.*, vol. 21, no. 7, pp. 526–534, 1973.

[10] P. Depalle, G. Garcia, and X. Rodet, "Tracking of partials for additive sound synthesis using hidden Markov models," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, vol. I, 1993, pp. 242–245.

[11] M. Day and S. Godsill, "Bayesian harmonic models for musical signal analysis," in *Proc. Seventh Valencia International meeting (Bayesian Statistics 7)*, 2002.

[12] S. Peleg, B. Porat, and B. Friedlander, "The achievable accuracy in estimating the instantanous phase and frequency of a constant amplitude signal," *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2216–2224, 1993.

[13] C. de Boor, *A Practical Guide to Splines*. New York: Springer-Verlag, 1978.

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

[15] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.

[16] W. H. Press, B. P. Flamery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipies, C-Version*. Cambridge University Press, 1989.

[17] A. Röbel, "A new approach to transient processing in the phase vocoder," in *Proc. of the 6th Int. Conf. on Digital Audio Effects (DAFx03)*, 2003, pp. 344–349.

[18] L. Cohen, *Time-frequency analysis*, ser. Signal Processing Series. Prentice Hall, 1995.

[19] S. Levine and J. O. Smith, "A sines+transients+noise audio representation for data compression and time/pitch-scale modifications," in *105th AES Convention*, 1998, preprint 4781.

[20] S. Levine, "Audio representations for data compression and compressed domain processing," Ph.D. dissertation, Department of Electrical Engineering, CCRMA, Stanford University, 1999.

[21] K. Fitz, L. Haken, and P. Christensen, "Transient preservation under transformation in an additive sound model," in *Proc. of the Int. Computer Music Conference (ICMC))*, 2000, pp. 392–395.

[22] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995.

[23] V. Rioux and M. Poletti, "An experimental sdif-sampler," in *Proc. of the International Computer Music Conference (ICMC)*, 2002, pp. 505–508.

[24] M. Wright, M. J. Beauchamp, K. Fitz, X. Rodet, A. Röbel, X. Serra, and G. Wakefield, "Analysis/synthesis comparison," *Organised Sound*, vol. 5, no. 3, pp. 173–189, 2000.

[25] M. Wright, "Sounds of the icmc2000 analysis/synthesis comparison session," 2000. [Online]. Available: http://cnmat.cnmat.berkeley.edu/SDIF/ICMC2000/sounds.html

[26] A. Röbel, "Adaptive additive synthesis using spline based parameter trajectory models," in *Proc. Int. Computer Music Conference (ICMC)*, 2001, pp. 369–372.

[27] R. M. Gray, "Toeplitz and circulant matrices: A review," Inform. Sys. Lab, Stanford University, Tech. Rep., 2002. [Online]. Available: http://ee.stanford.edu/ gray/toeplitz.pdf

APPENDIX

To determine the frequency resolution of the adaptive algorithm we study the impact of a stationary perturbing cosine on the optimal model parameters for a single nonstationary sinusoidal component. Hence, the signal to be studied is

$$\hat{S}(n) = A(n)\cos(\Phi(n)) + \Delta\cos(w_s n). \tag{22}$$

The parameter trajectories of the target sinusoid $A(n)$ and $\Phi(n)$ are assumed to match the trajectory model according to eq. (3) such that the trajectories can be represented without error. Using this signal, the influence of the disturbing cosine on the optimal partial parameters of a model comprising a single partial

$$\hat{P}(n) = \hat{A}(n)\cos(\hat{\Phi}(n)) = (\sum_{k=1}^{M} A_k b_k(n))\cos(\sum_{i=1}^{M}\Phi_i b_i(n)). \tag{23}$$

can be studied as a function of the frequency of the disturbance w_s. Here b_i are the B-splines used for amplitude and phase trajectories, which for simplicity are supposed to be of the same order and A_l and Φ_l are the free parameters of the respective parameter trajectories. Note that the use of fixed frequency and amplitude for the disturbing cosine imposes no restriction for the application of the result because any time limited signal can be replaced with arbitrary small error by means of a superposition of stationary cosines.

The minimum of the squared error eq. (4) is found by means of setting the gradient of the quadratic form with respect to the free parameters to zero

$$0 = 2\sum_n (\hat{S}(n) - \hat{P}(n))\frac{\partial \hat{P}(n)}{\partial A_i} \quad \forall\, i \in \{1, 2, \ldots, M\} \tag{24}$$

$$0 = 2\sum_n (\hat{S}(n) - \hat{P}(n))\frac{\partial \hat{P}(n)}{\partial \Phi_i} \quad \forall\, i \in \{1, 2, \ldots, M\} \tag{25}$$

For $\Delta = 0$ the global minimum is achieved if $\hat{A}(n) = A(n)$ and $\hat{\Phi}(n) = \Phi(n)$. We now want to establish the relation between the parameters A_i and Φ_i and $\Delta \neq 0$. To simplify the nonlinear relations between Φ_i and A_i and the amplitude of the perturbing signal and to obtain a fundamental idea about these relations we are linearizing eq. (24) and (25) around the optimum for $\Delta = 0$. The linearized equations will approximately describe the change of A_i and Φ_i for $\Delta \neq 0$ as long as the changes are sufficiently small. When linearizing the conditions in eq. (24) and (25), we perform all differentiations, make use of the fact that the error model error for $\Delta = 0$ is zero and that the bandwidth of the amplitude trajectory is small (eq. (13)). That leads us to the two separated linear systems

$$\begin{aligned}
&\Delta\sum_n \cos(w_s n) b_i(n)\cos(\Phi(n)) \\
&\quad\approx \sum_k \Delta_{A_k}\sum_n b_i(n)b_k(n)\cos^2(\Phi(n)) \\
&\quad\approx \frac{1}{2}\sum_k \Delta_{A_k}\sum_n b_i(n)b_k(n) \quad \forall\, i \in \{1, 2, \ldots, M\},
\end{aligned} \tag{26}$$

$$\begin{aligned}
&\Delta\sum_n \cos(w_s n) A(n) b_i(n)\sin(\Phi(n)) \\
&\quad\approx -\sum_k \Delta_{\Phi_k}\sum_n A(n)^2 b_i(n)b_k(n)\cos^2(\Phi(n))) \\
&\quad\approx \frac{-1}{2}\sum_k \Delta_{\Phi_k}\sum_n A(n)^2 b_i(n)b_k(n) \\
&\qquad\qquad \forall\, i \in \{1, 2, \ldots, M\},
\end{aligned} \tag{27}$$

that approximately describe the relations between the perturbing signal and the changes in the amplitude and phase parameter vectors denoted as Δ_{A_k} and Δ_{Φ_k}. The coefficient

matrices C_A and C_Φ that are related to the linear amplitude and phase equations in eq. (26) and eq. (27) have band-diagonal form with the coefficients c_{ik} given by the crosscorrelation between the B-splines $b_i(n)$ and $b_k(n)$. For C_Φ this correlation is additionally weighted by means of the amplitude trajectory of the target sinusoid. Accordingly, only the $(2o-1)$ inner diagonals are non zero and are monotonically decreasing with the distance from the main diagonal. The solution of these equations can be obtained by means of inversion of the coefficient matrices. Using notation $\{\Delta_{A_i}\}_i$ to represent a column vector with elements Δ_{A_i} we obtain

$$\{\Delta_{A_i}\}_i = C_A^{-1}\Delta\{\sum_n \cos(w_s n) b_k(n)\cos(\Phi(n))\}_k \quad (28)$$

$$\{\Delta_{\Phi_i}\}_i = C_\Phi^{-1}\Delta\{\sum_n \cos(w_s n) b_k(n) A(n)\sin(\Phi(n))\}_k. \quad (29)$$

For every parameter we may describe the impact of the perturbation by means of a superposition of injected errors E_{A_k} respectively E_{Φ_k} given by the three respectively four term products

$$E_{A_k} = \sum_n \cos(w_s n) b_k(n)\cos(\Phi(n)) \quad (30)$$

$$E_{\Phi_k} = \sum_n \cos(w_s n) b_k(n) A(n)\sin(\Phi(n)). \quad (31)$$

These equations can be interpreted as the real part of the Fourier transform of a signal derived from the target sinusoid that is evaluated at frequency w_s and uses the B-spline $b_k(n)$ as analysis window. In eq. (30) the transformation is applied to the target sinusoid having the amplitude set to be constant 1. In eq. (31) the transformation is applied to the target sinusoid after shifting its phase by $\frac{\pi}{2}$. Because $b_k(n)$ is used as analysis window its spectrum defines how the injected error depends on the frequency distance between the perturbing signal and the target sinusoid. We may conclude that the B-spline spectra define the frequency resolution of the adaptive algorithm.

The weighting factors for superimposing the injected errors to obtain the parameter change Δ_{A_i} and Δ_{Φ_i} are given by the coefficients in the i-th row of the inverted coefficient matrices. Due to the effects at the borders of the parameter trajectories and the fact that the target amplitude affects the coefficent matrices it is difficult to give general analytic expressions for these coefficients. Some insight can be obtained, however, if we restrict the analysis to the case of constant amplitude $(A(n) = A_0)$ and suppose that the target sinusoid is infinitely long. According to the theory described in [27] we can then approximate the inverted coefficent matrices by means of circular matrices that can be calculated from a single row in the center of the infinite coefficient matrix. The sequence of coefficients of the circular inverted matrix is then given by means of the inverse Fourier transform of the inverted Fourier transform of the original sequence of coefficients in the center row of the infinite coefficient matrices C_A and C_Φ. The resulting sequences that approximate the coefficients in the center of the matrices C_A^{-1} and C_Φ^{-1} for different polynomial orders are depicted in fig. 7. As expected the impact of injected error decreases with the distance between the position of the

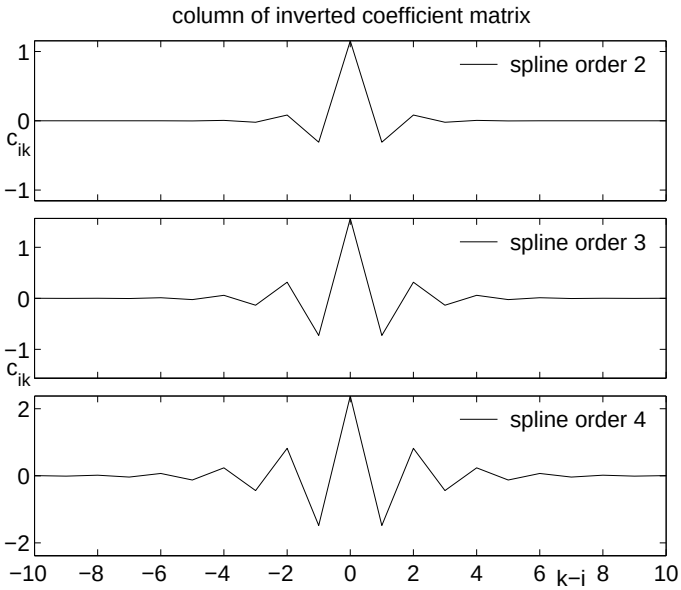

Fig. 7. Approximate coefficients c_{ik} of the column vector k of the inverted coefficient matrix for different polynomial orders and matrix size $N \to \infty$. The coefficients are maximal on the main diagonal at row i with $k - i = 0$.

B-spline coefficient i and the location of the injected error k. According to fig. 7 the injected error at position k is significant for the B-spline coefficients in a neighborhood of at least $\pm o$ segments. Note that the results related to the spread of perturbing energy hold similarly true when the frequency distance is 0. This is the case when modeling transients, where the perturbing signal is part of the sinusoid to be modeled, however, due to model insufficiencies cannot be expressed within the model.

Due to the linear approximation the results obtained above are valid only in case of small deviations from the optimal parameter vector. For most situations, however, the sidelobes of the B-spline are sufficiently small such that the impact of a perturbing signal outside the frequency range of the mainlobe can be described using the above relations.

PLACE PHOTO HERE

Axel Röbel received the Diploma in electrical engineering from Hannover University in 1990 and the Ph.D. degree (summa cum laude) in computer science from the Technical University of Berlin in 1993. In 1994 he joined the German National Research Center for Information Technology (GMD-First) in Berlin where he continued his research on adaptive modeling of time series of nonlinear dynamical systems. In 1996 he became assistant professor for digital signal processing in the communication science department of the Technical University of Berlin. In 2000 he obtained a research scholarship at CCRMA, Standford University, where he started the investigation into adaptive sinusoidal modeling. In 2000 he joined the analysis-synthesis team of IRCAM where he is currently doing research and development in the area of music and speech signal processing.