

Frequency-Zooming ARMA Modeling for Analysis of Noisy String Instrument Tones

Paulo A. A. Esquef

*Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000, FIN-02015 HUT, Espoo, Finland
Email: paulo.esquef@hut.fi*

Matti Karjalainen

*Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000, FIN-02015 HUT, Espoo, Finland
Email: matti.karjalainen@hut.fi*

Vesa Välimäki

*Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000, FIN-02015 HUT, Espoo, Finland
Pori School of Technology and Economics, Tampere University of Technology, P.O. Box 300, FIN-28101 Pori, Finland
Email: vesa.valimaki@hut.fi*

Received 31 May 2002 and in revised form 5 March 2003

This paper addresses model-based analysis of string instrument sounds. In particular, it reviews the application of autoregressive (AR) modeling to sound analysis/synthesis purposes. Moreover, a frequency-zooming autoregressive moving average (FZ-ARMA) modeling scheme is described. The performance of the FZ-ARMA method on modeling the modal behavior of isolated groups of resonance frequencies is evaluated for both synthetic and real string instrument tones immersed in background noise. We demonstrate that the FZ-ARMA modeling is a robust tool to estimate the decay time and frequency of partials of noisy tones. Finally, we discuss the use of the method in synthesis of string instrument sounds.

Keywords and phrases: acoustic signal processing, spectral analysis, computer music, sound synthesis, digital waveguide.

1. INTRODUCTION

It has been known for quite a long time that a free vibrating body may generate a sound that is composed of damped sinusoids, assuming valid the hypothesis of small perturbations and linear elasticity [1]. This behavior has motivated the use of a set of controllable sinusoidal oscillators to artificially emulate the sound of musical instruments [2, 3, 4]. As for analysis purposes, tools like the short-time Fourier transform (STFT) [5] and discrete cosine transform (DCT) [6] have been widely employed since these transformations are based on projecting the input signal onto an orthogonal basis consisting of sine or cosine functions.

An appealing idea, which is also based on resonant behavior of vibrating structures, consists in letting the resonant behavior be parametrically modeled by means of resonant filters (all-pole or pole-zero) excited by a source signal. For short duration excitation signals and filters parameterized by a few coefficients, such a source-filter model implies a compact representation for sound sources. Furthermore, para-

metric modeling of linear and time-invariant systems finds applications in several areas of engineering and digital signal processing, such as system identification [7], equalization [8], and spectrum estimation [9]. The moving-average (MA), the autoregressive (AR), and autoregressive moving-average (ARMA) models are among the most widely used ones. Indeed, there exists an extensive literature on estimation of these models [9, 10, 11, 12].

There is a long tradition in applying source-filter schemes in sound synthesis. For instance, the linear predictive coding (LPC) [13] used for speech coding and synthesis is one of the most well-known applications of source-filter synthesis. The problems involved in source-filter approaches can be roughly divided into two subproblems: the estimation of the filter parameters and the choice or design of suitable excitation signals. As regards the filter parameter estimation, standard techniques for estimation of AR and ARMA processes can be used. Ways of obtaining adequate excitations for the generator filter have been discussed in [14, 15, 16].

Model-based spectral analysis of recorded instrument sounds also finds applications in parametric sound synthesis. In this context, it is possible to derive the frequencies and decay times of the partial modes from the parameters of the estimated models (all-pole or pole-zero filters). This information can be used afterward to calibrate a synthesis algorithm, for example, a guitar synthesizer based on the computed waveguide method [17, 18].

However, when dealing with signals exhibiting a large number of mode frequencies, for example, low-pitched harmonic tones, high-order models are needed for properly modeling the signal resonances. Therefore, it is plausible to expect difficulties to either estimate or realize such high-order models.

A possible way to alleviate the burden of employing high-order models is to split the original frequency band into subbands with reduced bandwidth. Frequency-selective schemes allow signal modeling within a subband of interest with lower-order filters [14, 19, 20, 21]. Naturally, the choices of the subband bandwidth as well as the modeling orders depend on the problem at hand. For instance, in [20], Laroche shows that adequate modeling of beating modes of a single partial of a piano tone can be accomplished by applying a high-resolution spectral analysis method to the signal associated with the sole contribution of the specific partial. In this case, the decimated subband signal associated with the partial contribution was analyzed via the ESPRIT method [22].

In this paper, we review a frequency-zooming ARMA (FZ-ARMA) modeling technique that was presented in [23] and discuss the advantages of applying the method for analysis of string instrument sounds. Our focus, however, is not on the FZ-ARMA modeling formulation, which bears similarities to other subband modeling approaches, such as those proposed in [14, 20, 24, 25], among others. In fact, we are more interested in reliable ways to estimate the frequencies and decay times of partial modes when the tone under study is corrupted with broadband background noise. Within this scenario, our aim is to investigate the performance of the FZ-ARMA modeling as a spectrum analysis tool.

Every measurement setup is prone to noise interference to some extent, even in controlled conditions as in an anechoic environment. For instance, the recording circuitry involving microphones and amplifiers is one of the sources of noise. In [26], the authors highlight the importance of taking into account the level of background noise in the signal when attempting to estimate the decay time of string tone partials, especially for the fast decaying ones.

Another situation in which corrupting noise has to be carefully considered is in the context of audio restoration. In a recent paper [27], the authors proposed a sound source modeling approach to bandwidth extension of guitar tones. The method was applied to recover the high-frequency content of a strongly de-hissed guitar tone. To perform this task, a digital waveguide (DWG) model for the vibrating string has to be designed. In [27], the DWG model was estimated using a clean guitar tone similar to the noisy one. This resource was adopted because the presence of the corrupting noise prevented obtaining reliable estimates for the decay time of

high-frequency partials. These estimates were determined via a linear fitting over the time evolution of the partial amplitude (in dB), which was obtained through a procedure similar to the McAulay and Quatieri analysis scheme [2, 28].

Through examples which feature noisy versions of both synthetic and real string tones, we demonstrate that the FZ-ARMA modeling offers a reliable means to overcome the limitations of the STFT-based methods regarding estimating the decay time of partials.

This paper is organized as follows. Section 2 reviews the basic properties of AR and ARMA modeling and discusses signal modeling strategies in full bandwidth as well as in subbands. In Section 3, we formulate the FZ-ARMA modeling scheme and address issues related to the choice of the processing parameters. In Section 4, we employ the FZ-ARMA modeling to focus the analysis on isolated partials of synthetic and real string tones. Moreover, we assess the FZ-ARMA modeling performance on estimating the decay times of the partial modes under noisy conditions. In addition, we confront the results of spectral analysis of the subband signals using ARMA models against those obtained through the ESPRIT method. Section 5 discusses applications of the FZ-ARMA modeling in sound synthesis. In particular, we show an example in which, from the FZ-ARMA analysis of a noisy guitar tone, a DWG-based guitar tone synthesizer is calibrated. Conclusions are drawn in Section 6.

2. AR/ARMA MODELING OF STRING INSTRUMENT SOUNDS

2.1. Basic definitions

An ARMA process of order p and q , here indicated as $\text{ARMA}(p, q)$, can be generated by filtering a white noise sequence $e(n)$ through a causal linear shift-invariant and stable filter with transfer function [12]

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^q b_q(k)z^{-k}}{1 - \sum_{k=1}^p a_p(k)z^{-k}}. \quad (1)$$

For real-valued filter coefficients, the transfer function of an $\text{ARMA}(p, q)$ model has p poles and q zeros. Considering a flat power spectrum for the input, that is, $P_e(z) = \sigma_e^2$, the resulting output $x(n)$ has power spectrum given by

$$P_x(z) = \sigma_e^2 \frac{B_q(z)B_q^*(1/z^*)}{A_p(z)A_p^*(1/z^*)}, \quad (2)$$

where the symbol $*$ stands for complex conjugate.

An AR process is a particular case of an ARMA process when $q = 0$. Thus, the generator filter assumes the form

$$H(z) = \frac{b(0)}{1 - \sum_{k=1}^p a_p(k)z^{-k}}, \quad (3)$$

which is usually referred to as the transfer function of an all-pole filter.

2.2. Parameter estimation of AR and ARMA processes

Thorough descriptions of methods for estimation of AR and ARMA models are outside the scope of this paper since this topic is well covered elsewhere [9, 12] and computer-aid tools are readily available for this purpose. Here, we briefly summarize the most commonly used methods.

Parameter estimation of AR processes can be done by several means, usually through the minimization of a modeling error cost function. Solving for the model coefficients from the so-called autocorrelation and covariance normal equations [9] are perhaps the most common ways.

The stability of the estimated AR models is an important issue in synthesis applications. The autocorrelation method guarantees AR model estimates that are minimum phase. The Matlab function `ar.m` allows estimating AR models using several approaches [29].

Parameter estimation of ARMA processes is more complicated since the normal equations are no longer linear in the pole-zero filter coefficients. Therefore, the estimation relies on nonlinear optimization procedures that have to be done in an iterative manner. Prony's method and the Steiglitz-McBride iteration [30, 31] are examples of such schemes. A drawback of these methods is that the estimated pole-zero filters cannot be guaranteed to be minimum phase. In addition, and especially for high-order models, the estimated filters can be unstable. The functions `prony.m` and `stmcbr.m` are available in Matlab for estimation of ARMA models using Prony's and Steiglitz-McBride methods, respectively [32].

2.3. Full bandwidth modeling

Modeling of string instrument sounds has been approached by either physically motivated or signal modeling methods. Examples of the former can be found in physics-based algorithms for sound synthesis [18, 33, 34, 35]. Examples of the latter include the AR-based modeling of percussive sounds presented in [14, 15, 16, 36, 37].

In principle, when approaching the problem from a signal modeling point of view, it seems natural to employ a resonant filter, such as an all-pole or pole-zero filter, to model the mode behavior of a freely vibrating string, which consists of a sum of exponentially decaying sinusoids. However, modeling of broadband signals can be a tricky task. One practical issue related to both AR and ARMA modeling is model order selection. In general, there is no automated way to choose an appropriate order for the model assigned to a signal. For instance, one can deduce that AR modeling of low-pitched tones in full bandwidth is expected to require high-order models. The same is valid for piano tones which are produced by one to three strings sounding together. In this case, considering the detuning among the strings and two polarizations of transversal vibration per string, up to 6 resonance modes should be allocated to each partial of the tone.

In fact, the temporal envelope exhibited by partials of guitar and piano tones can be far from being exponentially decaying. On the contrary, the usually observed temporal envelopes contain frequency beating and two-stage decay [38].

This indicates that the partials are composed of two or more modes that are tightly clustered in frequency. The need for high-resolution frequency analysis tools is evident in these cases.

If frequency analysis is to be performed by means of AR/ARMA modeling, higher spectral resolutions can be attained by increasing the model orders. However, parameter estimation of high-order AR/ARMA models may be problematic if the poles of the system are very close to the unit circle and if there are poles located close to each other. Realizing a filter with these features is very demanding as the required dynamic range for the filter coefficients tends to be huge. In addition, computation of the roots associated with the corresponding polynomial in z , if necessary, can be also demanding and prone to numerical errors [39].

2.4. Frequency-selective modeling

The aforementioned problems have motivated the use of alternative modeling or analysis strategies based on subband decomposition [40]. In such schemes, the original signal is first split in several spectral subbands. Then, modeling or analysis of the resulting subband signals can be performed separately in each subband. Examples of subband modeling approaches can be found in [14, 16, 20, 24, 25].

A prompt advantage of subband decomposition of an AR/ARMA process is the possibility to focus the analysis on thinner portions of the spectrum. Thus, a small number of resonances can be analyzed at a time. This accounts for using lower-order models to analyze subband signals. Moreover, the subband signals can be down sampled, as their bandwidth is reduced compared to that of the original signal. As a consequence, the implied decrease in temporal resolution due to down-sampling is rewarded by an increase in frequency resolution. This fact favors the problem of resolving resonant modes that are very close to each other in frequency. The effects of decimating AR and ARMA processes have been discussed in [21, 41, 42].

3. FREQUENCY-ZOOMING ARMA METHOD

As presented in [23], the FZ-ARMA analysis consists of the following steps.

- (i) Define a frequency range of interest (for instance, to select a certain frequency region around the spectral peaks one wants to analyze).
- (ii) Modulate the target signal (shift in frequency by multiplying with a complex exponential) to place the center of the previously defined frequency band at the origin of the frequency axis.
- (iii) Lowpass filter the complex-valued modulated signal in order to attenuate its spectral content outside the band of interest.
- (iv) Down sample the lowpass filtered signal according to its new bandwidth.
- (v) Estimate an ARMA model for the previously obtained decimated signal. Throughout all examples shown in this work, the Steiglitz-McBride iteration method

[12, 30, 31] is employed to perform this task. More specifically, we used the `stmc_b.m` function available in the signal processing toolbox of Matlab [32].

In mathematical terms, and starting with a target sound signal $h(n)$, the first two steps of the FZ-ARMA method imply defining a modulation frequency f_m (in Hz) and multiplying $h(n)$ by a complex exponential, as to obtain the modulated response

$$h_m(n) = e^{-j\Omega_m n} h(n), \quad (4)$$

where $\Omega_m = 2\pi f_m / f_s$ with f_s being the sample rate. This modulation implies only a clockwise rotation of the poles of a hypothetical transfer function $H(z)$ associated with the AR process $h(n)$. Thus, if z_i is a pole of $H(z)$ with phase $\arg(z_i) = \Omega_i$, its resulting phase after rotation becomes

$$\Omega_{i,\text{rot}} = \Omega_i - \Omega_m. \quad (5)$$

The lowpass filtering is supposed to retain without distortion those poles located inside its passband. On the other hand, down sampling the resulting lowpass filtered response yields modified poles

$$z_{i,\text{zoom}} = z_i^{K_{\text{zoom}}} = |z_i|^{K_{\text{zoom}}} e^{j(\Omega_i - \Omega_m)K_{\text{zoom}}}, \quad (6)$$

where K_{zoom} is the zooming factor, which relates the new sampling rate to the original one as $f_{s,\text{zoom}} = f_s / K_{\text{zoom}}$.

Now, we know what the zooming procedure does to the poles, z_i , of the original transfer function. As a result, those poles, $\hat{z}_{i,\text{zoom}}$, estimated in subbands via ARMA modeling, need to be remapped to the original fullband domain. This can be accomplished by inverse scaling the poles and counter rotating them, that is,

$$\hat{z}_i = (\hat{z}_{i,\text{zoom}})^{1/K_{\text{zoom}}} e^{j\Omega_m}. \quad (7)$$

The frequency and decay time of the resonances present within the analyzed subband can be drawn from the angle and magnitude of \hat{z}_i , respectively.

Note that the original target response is supposed to be real valued and, therefore, its transfer function must have complex-conjugated pole pairs. However, due to the one-sided modulation performed in (4), the subband model returns pure complex poles. Thus, if the goal is to devise a real-valued all-pole filter in fullband for synthesizing the contribution of resonances within the analyzed subband, its transfer function must include not only the remapped poles, but also their corresponding complex-conjugates.

Hereafter, when referring to the models of the complex-valued subband signals, we will adopt the convention FZ-ARMA(p, q), where p and q stand for the orders of the denominator (AR part) and numerator (MA part), respectively.

3.1. Choice of parameters for the FZ-ARMA method

The choice of the FZ-ARMA parameters, that is, f_m , K_{zoom} , and the model orders, depends on several factors. We will now discuss these issues.

3.1.1. Zoom factor

Considering first the zoom factor, it can be said that the greater K_{zoom} , the higher the frequency resolution attainable in a subband. This favors cases in which the frequencies of the modes are densely clustered. However, large values of K_{zoom} imply a more demanding signal decimation procedure and shorter decimated signals.

The values of K_{zoom} and $f_{s,\text{zoom}}$ are tied together, and the latter defines the bandwidth of the subband which the analysis will be focused on. For instance, if the aim is to analyze the behavior of isolated partials of a tone, the choice of $f_{s,\text{zoom}}$ should be such that its value be less than two times the minimum frequency difference between adjacent partials. On the other hand, $f_{s,\text{zoom}}$ should be large enough to guarantee that the modes belonging to a given partial do not lie inside different subbands.

While the model estimation may be unnecessarily overloaded if based on long signals, it may yield poor results if based on few signal samples only. Therefore, the criterion upon which the value of $f_{s,\text{zoom}}$ is chosen should also take into account the number of samples that remains in the decimated signal.

3.1.2. Modulation frequency

Suppose that we are interested in analyzing a set of resonances concentrated around a frequency f_r . Having defined the bandwidth of the zoomed subband $f_{s,\text{zoom}}$, a straightforward choice is to set the value of the modulation frequency to $f_m = f_r$. Note that this option places the resonance peaks inside the subband around $\Omega_r = 0$. As pole estimation around $\Omega_r = 0$ may be more sensitive to numerical errors, we decided to adopt $f_m = f_r - f_{s,\text{zoom}}/8$, which implies concentrating the peaks around $\Omega_r = \pi/4$. This frequency shift is not harmful since the resonance peaks are still well inside the subband. Thus, their characteristics are not severely distorted by the nonideal lowpass filtering employed during the decimation procedure. However, to afford this choice of f_m and still ensure the isolation of a tone partial, the maximum value of $f_{s,\text{zoom}}$ should be at maximum one and half times the minimum frequency difference between adjacent partials.

The frequency of the partials can be predicted from that of the fundamental if the tone is harmonic or quasi-harmonic. However, as some level of dispersion is always present, errors at the frequencies of the higher partials are expected to occur. Alternatively, the frequencies of the partials can be determined by performing spectral analysis on the attack part of the tone and running a peak-picking algorithm over the resulting magnitude spectrum, as employed in [16, 25]. This approach is more general since it can deal with highly inharmonic tones.

In our experiments, we first estimate the fundamental frequency of the tone, a task that was performed through the multipitch estimator described in [43]. Then, after modeling the first partial, which allows obtaining a precise value of this partial frequency, the frequency of the following partial to be analyzed is set as the sum of the estimated frequency of the current partial with the value of the fundamental

frequency. This procedure is repeated until one reaches the desired number of partials to be analyzed. This approach minimizes the problems related to multiplicative errors when predicting the frequencies of higher partials based on integer multiples of the fundamental frequency.

3.1.3. Model order

Regarding the orders of the ARMA models, they should be chosen as to allow the modeling of the most prominent resonant modes of the signal. Depending on the case, *a priori* information on the characteristics of the signal at hand can be used to guide suitable model-order choices. For string instrument sounds, the estimation of the number of modes per partial can be based on the number of strings per note and the number of polarizations per string.

Moreover, it is known that if a real-valued signal has p resonant modes, one has to allocate at least two poles per resonant mode, that is, an ARMA($2p, 0$), to properly model it. However, due to the one-sided modulation used in the FZ-ARMA scheme, the resulting subband signals are complex valued, thus composed of pure complex poles. Therefore, only one single complex pole per mode suffices. As a consequence, at the expense of working with a complex arithmetic, the FZ-ARMA scheme optimizes the resources spent on modeling of the subband signals. This represents one advantage over, for instance, the modulation scheme proposed in [20], which yields real-valued decimated signals.

4. FZ-ARMA MODELING OF STRING INSTRUMENT TONES

In this section, we apply the FZ-ARMA modeling to analyze the resonant modes of isolated partials of string instrument sounds. We start by analyzing synthetic signals as a way to objectively evaluate the results. This allows knowing beforehand the mode frequencies and decay rates of the artificial tone. Thus, we can compare them with the estimates obtained via the FZ-ARMA modeling. In this context, the choice of the model orders is investigated as well as the modeling performance under noisy conditions. Then, following a similar analysis procedure, we evaluate the modeling performance of the FZ-ARMA method on recorded tones of real-world string instruments.

4.1. Experiments on artificially generated string instrument tones

4.1.1. Guitar tone synthesis

In this case study, the synthetic guitar tone is generated by means of a dual-polarization DWG model [18]. Thus, each of its partials has two modes with known parameters, that is, resonance frequencies and time constant of the exponentially decaying envelope.

The string model for one polarization is depicted in Figure 1. Its transfer function is given by

$$S(z) = \frac{1}{1 - z^{-L_i} H_{FD}(z) H_{LF}(z)}, \quad (8)$$

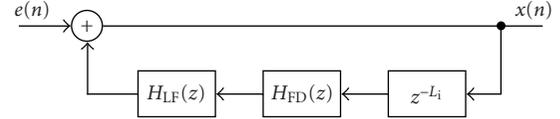


FIGURE 1: Block diagram of the string model.

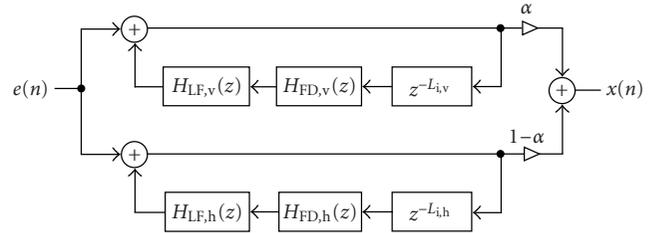


FIGURE 2: Block diagram of the dual-polarization string model. The subscripts “v” and “h” stand for vertical and horizontal, respectively.

where z^{-L_i} and $H_{FD}(z)$ are, respectively, the integer and fractional parts of the delay line associated with the length of the string. This length is given by $L = f_s/f_p$, where f_s and f_p are the sample frequency and fundamental frequency of the tone, respectively. The transfer function $H_{LF}(z)$ is called *loop filter* and is in charge of simulating the frequency-dependent losses of the partial modes.

For the sake of simplicity, we implemented the loop filter via the one-pole lowpass filter with transfer function given by

$$H_{LF}(z) = \frac{g(1+a)}{1+az^{-1}}. \quad (9)$$

The magnitude response of $H_{LF}(z)$ must not exceed unity in order to guarantee the stability of $S(z)$. This constraint imposes that $0 < g < 1$ and $-1 < a < 0$. As regards the fractional-delay filter $H_{FD}(z)$, we chose to employ the first-order allpass filter proposed in [44], which implies the computation of a single coefficient a_{fd} . This choice assures that the decay rates of the partials depend mainly on the characteristics of $H_{LF}(z)$.

The dual-polarization model consists in placing two string models in parallel as depicted in Figure 2. With this model, amplitude beating can be obtained by setting slightly different delay line lengths for each polarization. In addition, two-stage envelope decay can be accomplished by having loop filters with different magnitude responses for each polarization.

Consider first a string model with only one polarization. The partials of the resulting tone will decay exponentially and form a perfect harmonic series, that is, their frequencies are $f_v = \nu f_p$, where f_p is the fundamental frequency of the tone, and $\nu = 1, \dots, \lfloor f_s/(2f_p) \rfloor$ the partial indices. To determine the decay rate associated with each partial, we need to know the gain of the loop filter as well as the group delay of the feedback path (cascade of z^{-L_i} , $H_{FD}(z)$, and $H_{LF}(z)$) at the partial frequencies. By defining the partial frequencies

TABLE 1: Parameters used to generate the synthetic guitar tone. The sample rate was chosen as $f_s = 44.1$ kHz and α was set to 0.5.

Polarization	f_p	g	a	L_i	a_{fd}
Vertical	200 Hz	0.997	-0.03	220	0.3614
Horizontal	200.4 Hz	0.980	-0.10	219	0.0263

in radians as $\omega_\nu = 2\pi f_\nu/f_s$, the gain of the loop filter at ω_ν is given by

$$|H_{LF}(e^{j\omega_\nu})| = \frac{g(1+a)}{\sqrt{1+2a\cos(\omega_\nu)+a^2}}. \quad (10)$$

The group delay of a transfer function $F(z)$ is commonly defined as the ratio $\Gamma_F(\omega) = -\partial \arg\{F(e^{j\omega})\}/\partial\omega$. Then, if one defines $G(\omega_\nu)$ as the group delay (in samples) of the feedback path at ω_ν , that is, $G(\omega_\nu) = L_i + \Gamma_{H_{LF}}(\omega_\nu) + \Gamma_{H_{FD}}(\omega_\nu)$, the decay time (in seconds) of the partials can be obtained by

$$\tau_\nu = \frac{1}{f_s} \left(-\frac{G(\omega_\nu)}{\log(|H_{LF}(e^{j\omega_\nu})|)} \right). \quad (11)$$

Now we can generate an artificial guitar tone through the dual-polarization model, analyze it using the FZ-ARMA method, and compare the estimated values of the mode parameters with the theoretical ones. The tone is generated via the model shown in Figure 2 with parameters given in Table 1.

By adopting the parameters shown in Table 1, one guarantees that the modes of each partial will decay with different time constants. Hence, each partial exhibits a two-stage envelope decay behavior. Moreover, the mode frequencies of each partial are also different, thus yielding amplitude modulation in its envelope.

4.1.2. FZ-ARMA analysis

To proceed with the FZ-ARMA analysis of the generated tone, we have to choose appropriate values for the frequency bands of interest and corresponding modulation frequencies. In this example, equal bandwidth subbands are used to analyze the partials. The subband bandwidth is chosen to be equal to the fundamental frequency of the vertical polarization. This implies a new sampling frequency of $f_{s,zoom} = f_{p,v} = 200$ Hz for the subband signals and a zoom factor $K_{zoom} = 220$. For convenience, we only show results of parameter estimation up to the 45th partial. As highlighted in Section 3.1.2, for each partial frequency f_ν (of the vertical polarization) to be analyzed, the modulation frequency is chosen to be $f_m = f_\nu - f_{s,zoom}/8$.

The goal of this experiment is to gain an insight of the model orders that are necessary to reasonably estimate the mode parameters of the partials of a guitar tone. The FZ-ARMA procedure was devised in such a way that the subband signals are supposed to contain only two complex modes. Therefore, at least an FZ-ARMA(2, 0) must be employed to model each subband signal.

The results of mode parameter estimation obtained in this example are shown in Figure 3. Subplot 3(a) depicts

the reference values of the time constants of each polarization $\tau_{\nu,h}$ and $\tau_{\nu,v}$ as a function of the partial index ν . In subplots 3(c) and 3(e), one finds the relative errors in the time constant estimates, $\Delta\tau_\nu = |\tau_{\nu,ref} - \tau_{\nu,meas}|/\tau_{\nu,ref}$, when modeling the target signals through FZ-ARMA(2, 1) and FZ-ARMA(3, 2), respectively. Subplots 3(d) and 3(f) display the relative errors in the frequency estimates, $\Delta f_\nu = |f_{\nu,ref} - f_{\nu,meas}|/f_{\nu,ref}$, when modeling the target signals through FZ-ARMA(2, 1) and FZ-ARMA(3, 2), respectively.

From Figure 3, it is possible to verify that low-order models suffice to estimate the mode frequencies. On the contrary, to properly estimate the decay time of the partial modes, higher-order models are required. Furthermore, as one could expect, it is more difficult to estimate the time constants of faster decaying modes.

4.1.3. Analysis of noisy tones

We start with the same synthetic tone devised in Section 4.1.1. This tone is then corrupted with zero-mean white Gaussian noise, whose variance is adjusted to produce a certain signal-to-noise ratio (SNR) within the first 10 milliseconds of the tone. We proceed with the FZ-ARMA analysis of four noisy tones with SNR equal to 40, 20, 10, and 0 dB, respectively. The goal now is to investigate the effect of the SNR on the decay time estimates of the partial modes.

As in the previous example, equal-bandwidth subbands are used to analyze the partials of the tone. But, here, the adopted value of the zoom factor was $K_{zoom} = 600$. As before, the frequency f_ν of each partial to be analyzed defined the modulation frequency, which was chosen to be $f_m = f_\nu - f_{s,zoom}/8$. To model the two-mode partial signals, FZ-ARMA(3, 3) models were used. From the poles of each estimated model, those two with the largest radii were selected to determine the decay times and frequencies of the partial modes. In addition, for the sake of convenience, the estimated mode parameters were sorted by decreasing values of decay time.

The results are depicted in Figure 4, in which the solid and dashed lines describe the reference values of the decay time, associated with the vertical and horizontal polarizations, respectively, as functions of the partial indices. The circle and square markers indicate the corresponding estimated values.

As one could expect, the estimation performance is worsened when decreasing the SNR. Nevertheless, it is worth noting that even for the signal with SNR equal to 10 dB, the majority of the estimated values of decay time is concentrated around the reference values, especially for low-frequency partials. The occurring outliers can be either discarded, for example, negative values, or removed by means of median filtering. As for the mode frequency estimates (not shown), the maximum relative error encountered for the tone with SNR = 0 dB is of order equal to $\pm 0.1\%$, which is negligible.

4.1.4. Comparison against STFT-based methods

At this stage, one wonders if an estimation procedure based on short-time Fourier analysis or heterodyne filtering would

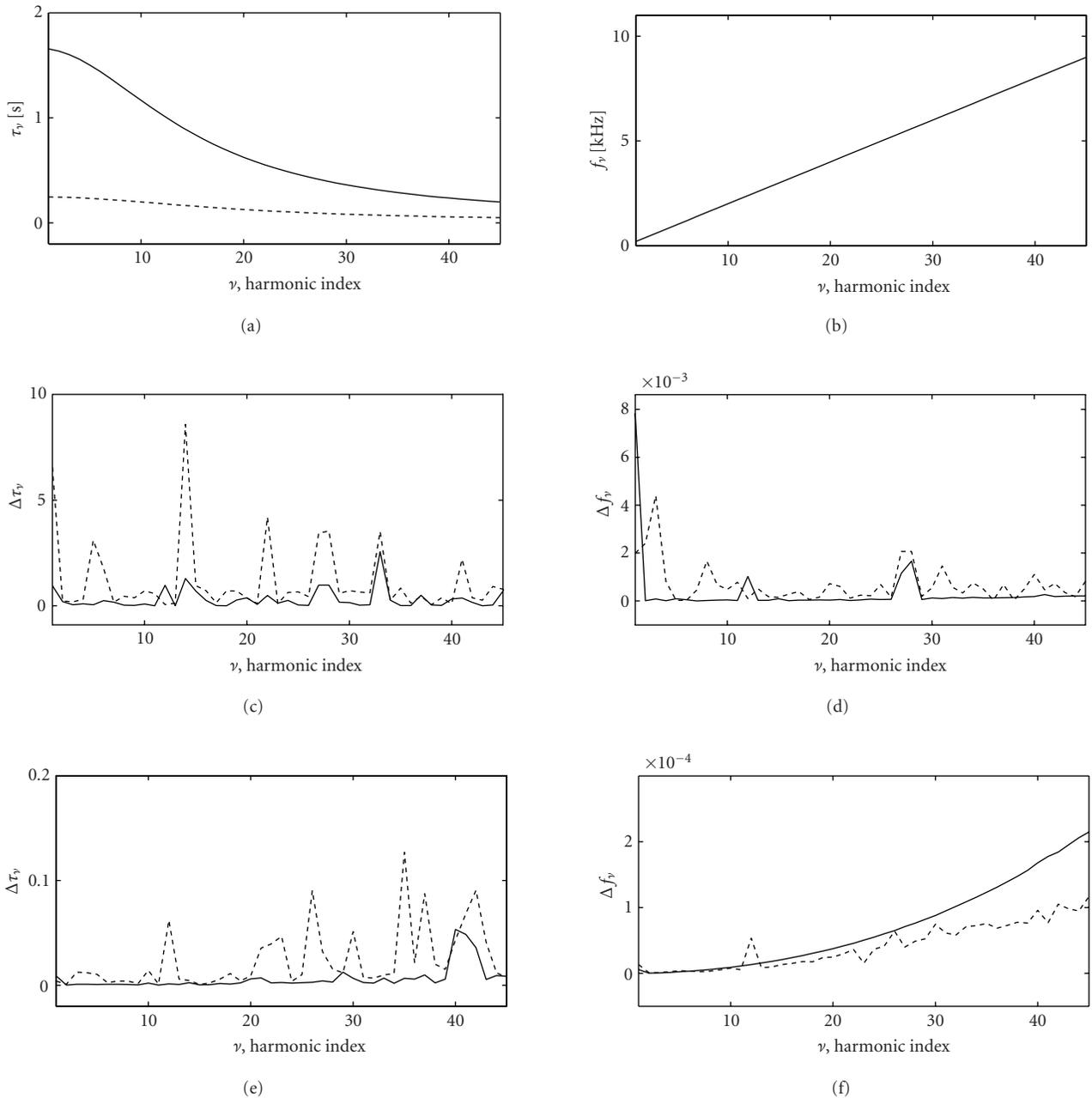


FIGURE 3: Case study on a synthetic string tone with amplitude envelope featuring beating and two-stage decay. Subplots (a) and (b) show, respectively, the reference time constants and frequencies of the modes as functions of the partial index; subplots (c) and (d) depict the relative errors $\Delta\tau_\nu = |\tau_{\nu,\text{ref}} - \tau_{\nu,\text{meas}}|/\tau_{\nu,\text{ref}}$ and $\Delta f_\nu = |f_{\nu,\text{ref}} - f_{\nu,\text{meas}}|/f_{\nu,\text{ref}}$ when estimating τ_ν and f_ν , respectively, via FZ-ARMA(2, 1) models; similar curves are shown in subplots (e) and (f) when adopting FZ-ARMA(3, 2) models. The results for the vertical and horizontal polarizations are indicated by solid and dashed lines, respectively.

yield similar results as those of the FZ-ARMA-based scheme when dealing with noisy signals.

In these approaches, each prominent partial is isolated somehow and the evolutions of its amplitude over time are tracked. Then, a linear slope is to be fitted to the obtained log-amplitude envelope curve. The decay time of the analyzed partial is determined from the slope of the fitted curve.

To start answering our question, we should remember that, even for clean signals, there are situations in which the just described slope fitting does not give appropriate results. Perhaps the most striking one is when the envelope curve shows amplitude beating. Back to the noisy signals, there may be a point in the amplitude envelope curves of the partials after which the noise component dominates the amplitude.

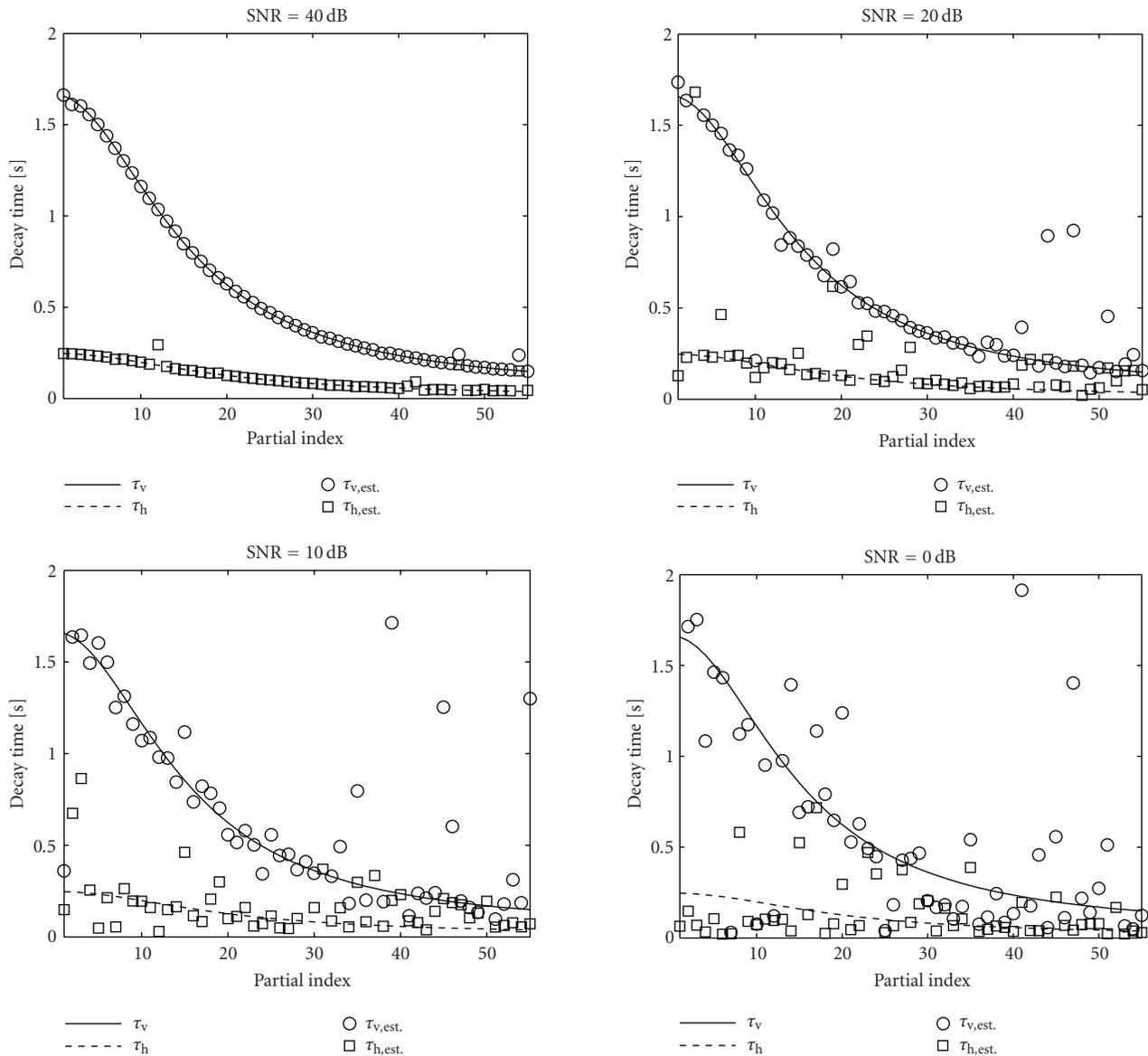


FIGURE 4: Decay times of two-mode partials of a synthetic noisy guitar tone: comparison between reference values and FZ-ARMA(3,3) estimates.

The noise floor is not so critical for the decay time estimation of low-frequency partials since they are usually stronger in amplitude and decay slowly. On the other hand, high-frequency partials are in general weaker in magnitude and decay fast. They are likely to reach and be masked by the noise floor very early in time. Taking into account the noise floor level is essential for the decay time estimation of these partials (see [26, Figure 5]).

For the sake of simplicity, we do not use neither the heterodyne filtering nor the sinusoidal modeling (SM) analysis in the comparisons shown in this section. Instead, we can resort to the frequency-zooming procedure itself. The amplitude envelope curves of each partial are obtained directly from the evolution of the signal magnitude within each

subband. Note that we are dealing with narrow subbands (bandwidth of about 70 Hz) and that each subband isolates a given partial. Therefore, the so-attained envelope curves will approximate well the curves that would result from either the heterodyne filtering or the SM analyses. The latter, however, would provide smoother curves. Yet, they would inevitably be lower-bounded by the average amplitude of the noise floor.

As an example, we compare the analysis of two high-frequency partials (6th and 13th) of the string tone devised in Section 4.1. These high-order partials are chosen on purpose to illustrate the effect of the corrupting noise on the amplitude envelope curves. Figure 5 compares the envelope curves of the featured partials in 3 conditions: noiseless tone

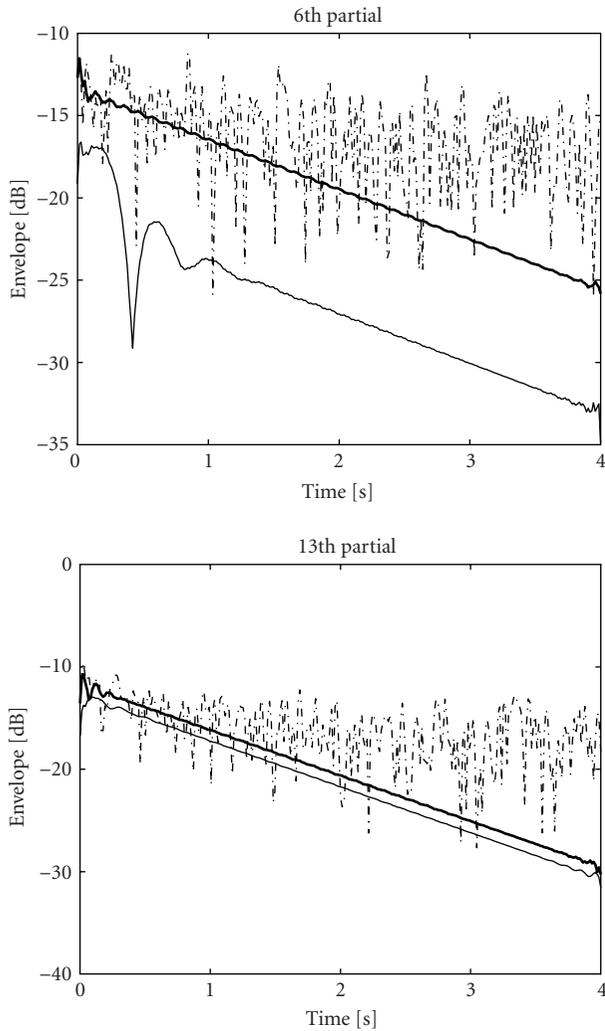


FIGURE 5: Analysis of the 6th and 13th partials of the synthetic tone: comparison among the envelopes of the reference signal (thinner solid line), its noisy version with $\text{SNR} = 0$ dB (dash-dotted line), and the modeled signal via FZ-ARMA(3, 3) (thicker solid line) based on the noisy signal.

(thinner solid line), noisy signal with $\text{SNR} = 0$ dB (dash-dotted line), and modeled signal based on the noisy target (thicker solid line).

From Figure 5, it becomes evident that, for the noisy signal, decay time estimation of the partials via slope fitting is impractical. On the contrary, the FZ-ARMA modeling is capable of properly estimating the decay time of the slowest decaying or the most prominent partial mode. Note that we are primarily interested in the slope of the envelope curve. The upward bias, which is observed in the envelopes of the modeled signals, occurs due to the difference in power between the clean and the noisy version of the signal.

The frequency-zooming procedure per se accounts for a significant improvement in the value of the SNR. For instance, if the target signal is a single complex exponential immersed in white noise, the reduction in SNR due to

the zooming will be given by $10 \log_{10}(K_{\text{zoom}})$. Of course, an even bigger SNR improvement can be achieved by FFT-based analysis. This comes from the fact that tracking a single frequency bin in the DFT domain (preferably refined by parabolic interpolation) implies analysis within a much narrower bandwidth than the frequency-zooming scheme. However, the improvement in the SNR is not the main issue here. This larger SNR improvement does not prevent the amplitude envelope from being lower-bounded by the noise floor level after some time.

The keypoint here is that fitting a parametric model to the partial signals allows capturing the intrinsic temporal structures of them, even in noise conditions. Moreover, the resonance features are derived from the model parameters rather than from a simple curve fitting process. As a consequence, a further improvement in the SNR is achieved, culminating in more reliable estimates for the decay time of the partials. Of course, the corrupting noise tends to degrade and bias the estimated models. Thus, any improvement in the SNR before the modeling stage is welcome. The frequency zooming helps in this matter as well.

4.1.5. Comparison against ESPRIT method

One could also think of applying other high-resolution spectral analysis methods to the subband signals. For instance, Laroche has used the ESPRIT method [20, 22] to analyze modes of isolated partials of clean piano tones. Just for comparison purposes, we repeat the experiments conducted in Section 4.1.3 using the ESPRIT method [22, 45]. More precisely, we employ the frequency-zooming procedure as before, but replace the ARMA modeling with the ESPRIT method as a means to analyze the subband signals.

In the ESPRIT method, we have to set basically three parameters: the length of the signal to be analyzed, N , the a priori estimate of the number of complex exponentials in the signal, M , and the pencil parameter, $M \leq P_{\text{pencil}} \leq N - M$. Analysis of noise sensitivity of the ESPRIT method has been conducted in [45] for single complex exponentials in noise. It revealed that setting $P_{\text{pencil}} = N/3$ or $P_{\text{pencil}} = 2N/3$ are the best choices for the pencil parameter, in order to minimize the effects of the noise on the exponential estimates. Furthermore, as highlighted in [20], overestimating M is harmless and even desirable to avoid biased frequency estimates. The ESPRIT method outputs M complex eigenvalues from which the frequency and decay time of M exponentials can be derived. As M is usually overestimated, a pruning scheme has to be employed to select the most prominent exponentials. In our experiments, we take only the two exponentials with the largest decay times.

According to the results of our simulations, the performances of the ESPRIT and ARMA methods are equivalent for estimating the frequencies of the resonant modes. For instance, as regards the frequency estimates, the maximum relative errors measured for the tone with $\text{SNR} = 0$ dB were 0.19 and 0.11, respectively, for the ESPRIT and ARMA methods. In this particular example, FZ-ARMA(3, 3) models were used whereas the parameter values adopted in the ESPRIT method were $N = 295$, $P_{\text{pencil}} = 98$, and $M = 20$.

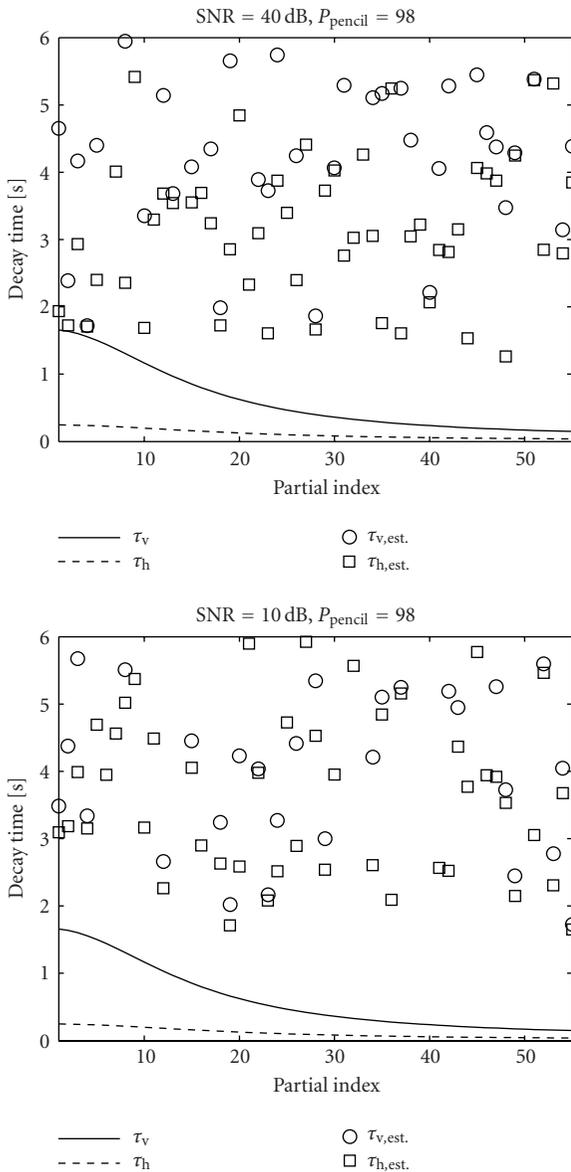


FIGURE 6: Decay times of partial modes of synthetic noisy guitar tones: comparison between reference values and ESPRIT estimates ($M = 20$ and $P_{\text{pencil}} = 98$).

The situation is different when it comes to the decay time estimates. It seems that the accuracy of these estimates is very dependent on the choice of pencil parameter. For instance, when dealing with noisy signals, setting $P_{\text{pencil}} = M$ yields underestimated values of decay time. On the contrary, increasing the value of P_{pencil} tends to produce overestimated values of decay times. According to the results of our experiments, this is also the case if $P_{\text{pencil}} = N/3$ is chosen.

Figure 6 confronts the reference values of the decay time against the estimates obtained through the ESPRIT method with $M = 20$ and $P_{\text{pencil}} = 98$. It can be clearly seen that the decay time estimates are substantially overestimated, even for moderate levels of SNR. Interestingly enough, repeating the

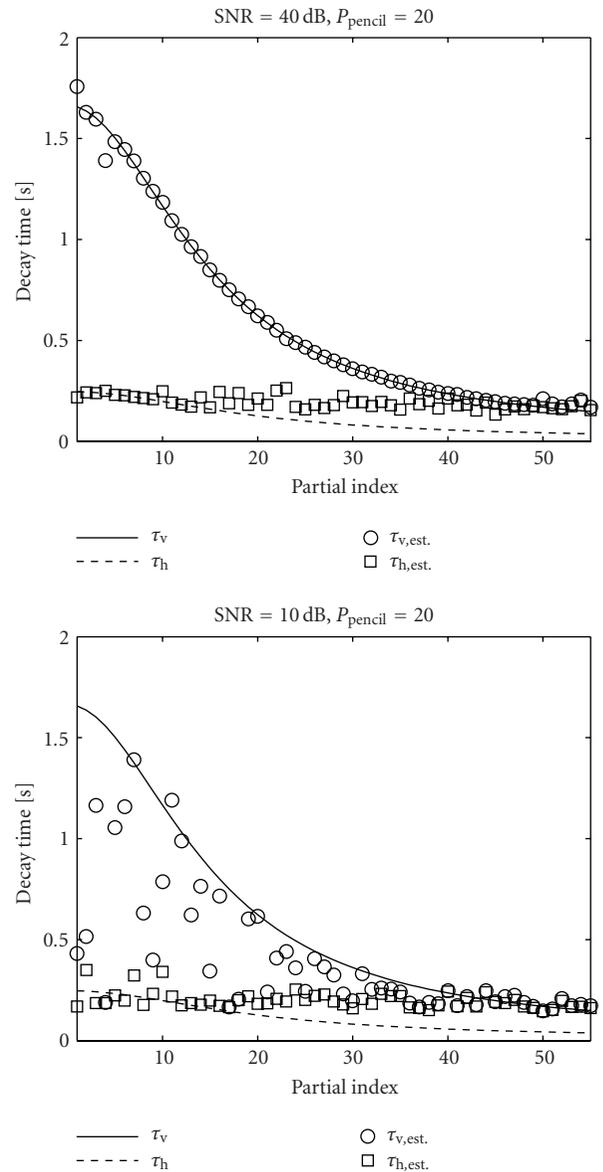


FIGURE 7: Decay times of partial modes of synthetic noisy guitar tones: comparison between reference values and ESPRIT estimates ($M = 20$ and $P_{\text{pencil}} = 20$).

experiments for $P_{\text{pencil}} = M = 20$ yields better results, as can be seen in Figure 7. In this case, the estimates are much more accurate than those obtained with $P_{\text{pencil}} = 98$. Notwithstanding, these estimates are still worse than those drawn from the poles of the ARMA(3, 3) fitted to the subband signals, as one can verify from Figure 4. Therefore, we stick to the FZ-ARMA modeling in the following experiments.

4.1.6. Discussion

Carrying out systematic performance comparisons among the addressed methods of decay time estimation is outside the scope of this work. Including such comparisons would demand not only covering a broader range of situations

and examples, but also precise description of the algorithms and the calibration of their associated processing parameters. Besides, comparisons between FFT-based schemes of spectral analysis, such as the SM technique, and parametric approaches are not fair. Sticking to comparisons among parametric methods of spectral analysis would necessarily include other techniques than just the ARMA and ESPRIT methods.

The comparisons shown in Section 4.1.4 are basically meant to highlight the situations in which STFT-based methods for decay time estimation are prone to failure. A presumed goal is to motivate the need for alternative solutions to decay time estimation in noisy conditions.

As for the performance comparisons between the ARMA and the ESPRIT methods, they were conducted after the frequency-zooming stage in order to keep equal conditions. Yet, the performance results can depend significantly on the choice of the processing parameters. This fact is clearly verified by comparing the results shown in Figures 6 and 7. Moreover, translating the parameters of one method into those of the other may not be straightforward. Due to the aforementioned reasons, we restrict the comparisons to a single case study. Rather than tabulating the attained performances, we believe that visual assessment on Figures 4, 6, and 7 offers more effective means of drawing conclusions on the results.

In summary, the STFT-based schemes are appropriate for decay time estimation of the partials when the partials show monotonic and exponential decay and when the measurement noise is low. If the noise component is prominent, reliable decay time (and frequency) estimation of the high-order partials will be prevented. For both the parametric methods tested, and under the setups adopted, a reliable frequency estimation for the partials of noisy tones is attained. As regards the decay time estimation in noisy conditions, the ARMA analysis performs better in general than the ESPRIT method.

Now, we comment specifically on the analysis results of the noisy tone with $\text{SNR} = 20$ dB. The ESPRIT method seems to overestimate the decay times as the value of the pencil parameter increases. Adopting the minimum value for the pencil parameters yielded the best results. Yet, the ESPRIT analysis underestimates the decay times of the low-order partials. This is critical from the perceptual point of view, especially if one aims at resynthesizing a new tone based on the analyzed data. For the high-order partials, however, the ESPRIT-based decay time estimates seem to converge with low variance to the decay time of the slowest resonance mode. In contrast, there are more outliers in the decay time estimates attained via the ARMA analysis. Nevertheless, the ARMA analysis seems to do a better job in properly segregating the estimates into two distinct resonance modes.

Finally, when it comes to choosing the most appropriate technique, many variables should be considered. Examples of such variables are the characteristics of the problem at hand and the aimed objectives, the effectiveness of the available tools in performing the targeted task, and the available computational resources. The latter issue, although important, does not fit to the profile of this paper. Therefore, discussions

on the computational complexity of the tested methods are not included.

4.2. Experiments on recorded string instrument tones

In this section, we follow the same methodology used in Sections 4.1.2 and 4.1.3 to analyze recorded tones of real-world string instruments. Here, we do not have a set of reference values for the decay times of the partials. Nevertheless, based on the results obtained for the synthetic tone, we can assume that the FZ-ARMA modeling of an originally clean tone provides correct estimates for the decay time of the partial modes. Then, this set of values can be taken as a reference.

For this experiment, we selected a clean classical guitar tone A2 ($f_p = 109.97$ Hz, softly plucked open 5th string), which was recorded in anechoic conditions. Three noisy versions of this tone, with $\text{SNR} = 60$, $\text{SNR} = 40$, and $\text{SNR} = 20$ dB, respectively, were generated by adding zero-mean white Gaussian noise to the clean tone. The noise variance was adjusted as to produce the desired SNR during the attack part of the tone (about 20 milliseconds starting from the maximum amplitude).

The first step of the analysis procedure is to obtain an estimate of the fundamental frequency of the noisy tone. This estimate is the starting point to the choices of the bandwidth of the subbands and the modulation frequencies to be used in the FZ-ARMA analysis. The fundamental frequency of tone with $\text{SNR} = 20$ dB was estimated to be $f_p = 110.25$ Hz, which is not far from that of the clean tone. Thus, by following the guidelines stated in Section 3.1.2, we can proceed toward analyzing the higher partials of both the clean and the noisy tones. The parameters used in the FZ-ARMA analysis were $K_{\text{zoom}} = 600$, $f_m = f_v - f_{s,\text{zoom}}/8$, and FZ-ARMA(3,3) models. This time, only the decay time of the slowest decaying mode of each partial was extracted.

The results of this experiment are displayed in Figure 8. The solid line curves correspond to the estimated values of decay time based on the original clean tone. On the other hand, the circles show the corresponding estimated values based on the noisy tones with indicated SNRs. From Figure 8, we observe that, even for the tone with $\text{SNR} = 20$ dB, the FZ-ARMA analysis provides reliable decay time estimates, especially for the low-frequency partials.

5. APPLICATIONS IN SOUND SYNTHESIS

5.1. Digital waveguide synthesis

We have seen in Section 4 that the FZ-ARMA modeling can be used as an analysis tool, aiming at estimating the parameters associated with the resonances of the tone partials. Thus, based on the set of frequencies and decay times estimated for each partial, one could design a DWG model to resynthesize the tone.

More interestingly, the FZ-ARMA modeling allows estimating more than one frequency and decay time per partial. Thus, one can consider using this information to design the filters of a multipolarization DWG model, such as the dual-polarization DWG model shown in Figure 2. As in source-

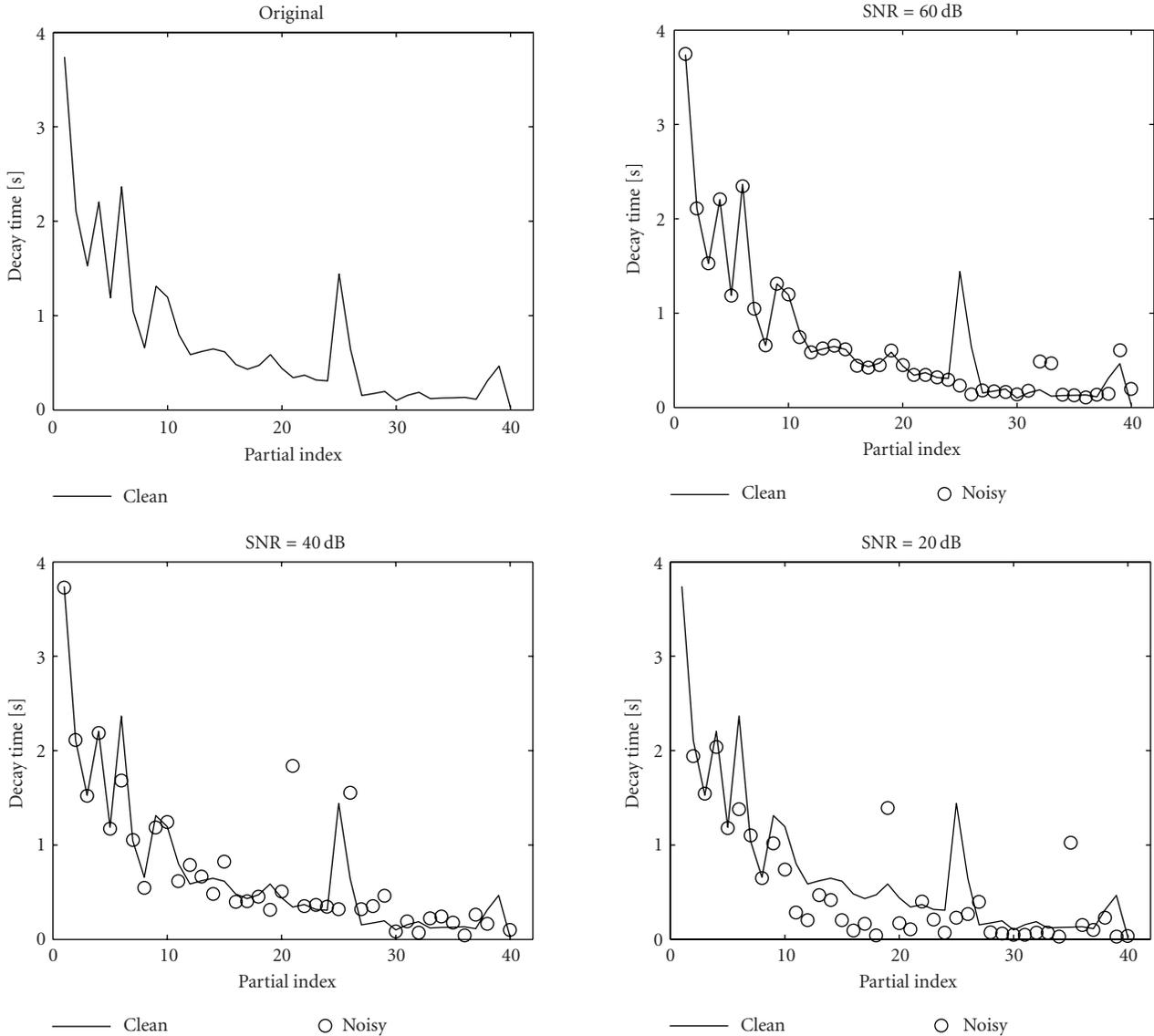


FIGURE 8: FZ-ARMA(3, 3) estimates of the decay time of partials of an A2 guitar tone: comparisons among estimates based on the original clean signal and its noisy versions at different SNRs.

filter synthesis, in DWG-based synthesis, the excitation signal is in charge of controlling the initial phase and amplitude of the resonance modes. In this work, however, we will not tackle the attainment of suitable excitation signals but we concentrate more on the calibration of the string models.

Calibrating a multipolarization DWG model based on the estimated parameters of the partial modes is a difficult task, especially when dealing with real-world recorded tones immersed in noise. This is mainly due to the high variance exhibited in the estimates of decay time of the partial modes. In contrast to what is seen in the analysis results of the synthetic tone shown in Section 4.1.2, the decay time of the partial modes, estimated from a recorded tone, cannot be easily discriminated in two or more distinct classes. Thus, deciding which partial mode belongs to which polarization turns out

to be a difficult nonlinear optimization problem. We leave this topic for future research and we stick to the calibration of the one-polarization DWG model.

5.1.1. Calibration of one-polarization DWG model from noisy tones

We start with an example in which the target signal is the corrupted version (SNR = 20 dB) of the recorded guitar tone featured in Section 4.2. From the FZ-ARMA analysis of this tone, we obtained estimates for the frequency and decay time of the partial modes. Then, the specification for the magnitude of the loop filter at the partial frequencies can be obtained by

$$|H_{LF}(f_v)| = e^{-\nu/f_v\tau_v}, \quad (12)$$

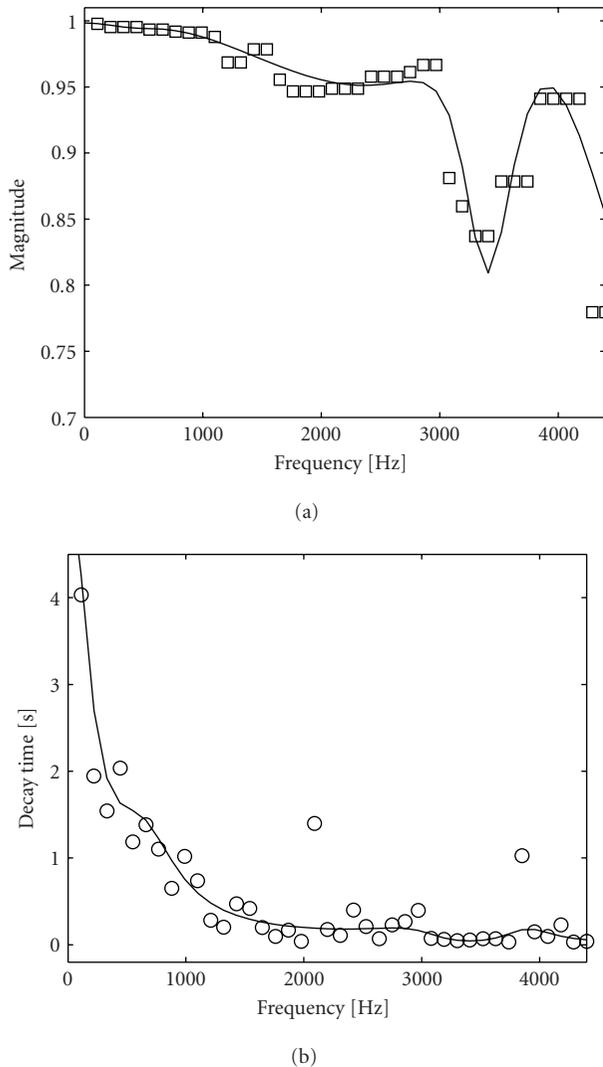


FIGURE 9: Specification points and attained response of the 8th-order IIR loop filter: (a) smoothed magnitude specification (squares) versus attained response (solid line) up to the frequency of the 40th partial; (b) measured decay times (circles) versus attained values forged by the loop filter response (solid line).

where ν is the partial index, f_ν are the frequencies of the partials in Hz, and τ_ν are the corresponding decay times in seconds.

As the sequence of estimated decay times, which was based on the corrupted signal, seems to have a couple of outliers, it was first median filtered using a three-sample window. The values of τ_ν that result from the filtered sequence are then used in (12).

The specification of the loop filter within the frequency range above the frequency of the 40th partial is devised artificially. We fit a -6 dB per octave slope to the magnitude specification points associated with the highest 10 partials and extrapolate the curve up to the Nyquist frequency. To design a loop filter that approximates this extended specification, we resort to the IIR design method proposed in [46, 47].

Figure 9 shows the results obtained by approximating the specified (smoothed) magnitude response of the loss filter via an 8th-order IIR lowpass filter.

We could also think of designing a dispersion filter for the DWG model. In this case, the specification for phase response of the allpass dispersion filter could be based on the estimated frequencies of the partials in a similar manner to what was done in [48, 49]. However, for the noisy tone under study, the variance observed in these estimates prevented one from obtaining any meaningful specification for the dispersion filter.

6. CONCLUSION

In this paper, a spectral analysis technique based on FZ-ARMA modeling was applied to string instrument tones. More specifically, the method was used to analyze the resonant characteristics of isolated partials of the tones. In addition, analyses performed on noisy tones demonstrated that the FZ-ARMA modeling turns out to be a robust tool for estimating the frequencies and decay times of the partial modes, despite the presence of the corrupting noise. Comparisons between the estimates attained by FZ-ARMA modeling and those obtained via the ESPRIT method revealed a superior performance of the former method when dealing with noisy tones. Finally, the paper discussed the use of FZ-ARMA modeling in sound synthesis. In particular, the calibration of a DWG guitar synthesizer was successfully carried out based on FZ-ARMA analysis of a recorded guitar tone, which was artificially corrupted by zero-mean white Gaussian noise.

ACKNOWLEDGMENTS

The work of Paulo A. A. Esquef has been supported by a scholarship from the Brazilian National Council for Scientific and Technological Development (CNPq-Brazil) and by the Academy of Finland project “Technology for Audio and Speech Processing.” The authors wish to thank Mr. Balázs Bank, Dr. Cumhur Erkut, and Dr. Lutz Trautmann for kindly providing some of the codes used in the simulations. Finally, the authors would like to thank the anonymous reviewers for their comments, which contributed to the improvement of the quality of this manuscript.

REFERENCES

- [1] A. H. Benade, *Fundamentals of Musical Acoustics*, Dover Publications, Mineola, NY, USA, 1990.
- [2] R. J. McAulay and T. F. Quatieri, “Speech analysis/synthesis based on a sinusoidal representation,” *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.
- [3] J. O. Smith III and X. Serra, “PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation,” in *Proc. International Computer Music Conference (ICMC ’87)*, Champaign-Urbana, Ill, USA, 1987.
- [4] J. C. Maher, “Sinewave additive synthesis revisited,” in *91st AES Convention*, New York, NY, USA, October 1991.
- [5] J. B. Allen and L. R. Rabiner, “A unified approach to short-

- time Fourier analysis and synthesis," *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1558–1564, 1977.
- [6] H. S. Malvar, *Signal Processing with Lapped Transforms*, Artech House, Norwood, Mass, USA, 1992.
- [7] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.
- [8] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 3rd edition, 1996.
- [9] S. M. Kay, *Modern Spectral Estimation*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.
- [10] A. V. Oppenheim, A. Willsky, and I. Young, *Signals and Systems*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1983.
- [11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.
- [12] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, New York, NY, USA, 1996.
- [13] J. Makhoul, "Linear prediction: a tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.
- [14] J. Laroche, "A new analysis/synthesis system of musical signals using Prony's method. Application to heavily damped percussive sounds," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 3, pp. 2053–2056, Glasgow, Scotland, UK, May 1989.
- [15] J. Laroche and J.-L. Meillier, "Multichannel excitation/filter modeling of percussive sounds with application to the piano," *IEEE Trans. Speech, and Audio Processing*, vol. 2, no. 2, pp. 329–344, 1994.
- [16] M. W. Macon, A. McCree, W.-M. Lai, and V. Viswanathan, "Efficient analysis/synthesis of percussion musical instrument sounds using an all-pole model," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 6, pp. 3589–3592, Seattle, Wash, USA, May 1998.
- [17] J. O. Smith III, "Efficient synthesis of stringed musical instruments," in *Proc. International Computer Music Conference (ICMC '93)*, pp. 64–71, Tokyo, Japan, September 1993.
- [18] M. Karjalainen, V. Välimäki, and Z. Jánosy, "Towards high-quality sound synthesis of the guitar and string instruments," in *Proc. International Computer Music Conference (ICMC '93)*, pp. 56–63, Tokyo, Japan, September 1993.
- [19] J. Makhoul, "Spectral linear prediction: Properties and applications," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 23, no. 3, pp. 283–296, 1975.
- [20] J. Laroche, "The use of the matrix pencil method for the spectrum analysis of musical signals," *Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 1958–1965, 1993.
- [21] L. W. P. Biscainho, P. S. R. Diniz, and P. A. A. Esquef, "ARMA processes in sub-bands with application to audio restoration," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 2, pp. 157–160, Sydney, Australia, May 2001.
- [22] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1340–1342, 1986.
- [23] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkitvirta, and V. Välimäki, "Frequency-zooming ARMA modeling of resonant and reverberant systems," *Journal of the Audio Engineering Society*, vol. 50, no. 12, pp. 1012–1029, 2002.
- [24] J. Laroche and J.-L. Meillier, "A simplified source/filter model for percussive sounds," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 173–176, New York, NY, USA, October 1993.
- [25] R. B. Sussman and M. Kahrs, "Analysis and resynthesis of musical instrument sounds using energy separation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, pp. 997–1000, Atlanta, Ga, USA, May 1996.
- [26] C. Erkut, V. Välimäki, M. Karjalainen, and M. Laurson, "Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar," in *108th AES Convention*, Paris, France, February 2000, preprint 5114. Available on-line at <http://lib.hut.fi/Diss/2002/isbn9512261901/>.
- [27] P. A. A. Esquef, V. Välimäki, and M. Karjalainen, "Restoration and enhancement of solo guitar recordings based on sound source modeling," *Journal of the Audio Engineering Society*, vol. 50, no. 4, pp. 227–236, 2002.
- [28] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy, "Physical modeling of plucked string instruments with application to real-time sound synthesis," *Journal of the Audio Engineering Society*, vol. 44, no. 5, pp. 331–353, 1996.
- [29] MathWorks, "MATLAB System Identification Toolbox," 2001, User's Guide.
- [30] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automatic Control*, vol. 10, no. 4, pp. 461–464, 1965.
- [31] K. Steiglitz, "On the simultaneous estimation of poles and zeros in speech analysis," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 25, no. 3, pp. 229–234, 1977.
- [32] MathWorks, "MATLAB Signal Processing Toolbox," 2001, User's Guide.
- [33] J. O. Smith III, *Techniques for digital filter design and system identification with application to the violin*, Ph.D. thesis, Elec. Eng. Dept., Stanford University, Stanford, Calif, USA, 1983.
- [34] J. O. Smith III, "Physical modeling using digital waveguides," *Computer Music Journal*, vol. 16, no. 4, pp. 74–91, 1992.
- [35] M. Karjalainen and J. O. Smith III, "Body modeling techniques for string instrument synthesis," in *Proc. International Computer Music Conference (ICMC '96)*, pp. 232–239, Hong Kong, China, August 1996.
- [36] M. Sandler, "Analysis and synthesis of atonal percussion using high order linear predictive coding," *Applied Acoustics*, vol. 30, no. 2-3, pp. 247–264, 1990.
- [37] J.-L. Meillier and A. Chaigne, "AR modeling of musical transients," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 3649–3652, Toronto, Canada, April 1991.
- [38] G. Weinreich, "Coupled piano strings," *Journal of the Acoustical Society of America*, vol. 62, no. 6, pp. 1474–1484, 1977.
- [39] M. Sandler, "Algorithm for high precision root finding from high order LPC models," *IEE Proceedings. Part I: Communications, Speech and Vision*, vol. 138, no. 6, pp. 596–602, 1991.
- [40] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.
- [41] K. B. Eom and R. Chellappa, "ARMA processes in multirate filter banks with applications to radar signal classification," in *Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 136–139, Philadelphia, Pa, USA, October 1994.
- [42] A. Benyassine and A. N. Akansu, "Subspectral modeling in filter banks," *IEEE Trans. Signal Processing*, vol. 43, no. 12, pp. 3050–3053, 1995.
- [43] T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Trans. Speech, and Audio Processing*, vol. 8, no. 6, pp. 708–716, 2000.
- [44] D. A. Jaffe and J. O. Smith III, "Extensions of the Karplus-Strong plucked-string algorithm," *Computer Music Journal*, vol. 7, no. 2, pp. 56–69, 1983.
- [45] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814–824, 1990.
- [46] B. Bank, "Physics-based sound synthesis of the piano," Tech. Rep. 54, Laboratory of Acoustics and Audio Signal Processing,

Helsinki University of Technology, Espoo, Finland, June 2000, available on-line at <http://www.acoustics.hut.fi/publications/>.

- [47] B. Bank and V. Välimäki, "Robust loss filter design for digital waveguide synthesis of string tones," *IEEE Signal Processing Letters*, vol. 10, no. 1, pp. 18–20, 2003.
- [48] D. Rocchesso and F. Scalcon, "Accurate dispersion simulation for piano strings," in *Proc. Nordic Acoustical Meeting (NAM '96)*, pp. 407–414, Helsinki, Finland, June 1996.
- [49] L. Trautmann, B. Bank, V. Välimäki, and R. Rabenstein, "Combining digital waveguide and functional transformation methods for physical modeling of musical instruments," in *Proc. AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, pp. 307–316, Espoo, Finland, June 2002.

Paulo A. A. Esquef was born in Brazil, in 1973. He received the Engineering degree from Polytechnic School of the Federal University of Rio de Janeiro (UFRJ) in 1997 and the M.S. degree from COPPE-UFRJ in 1999, both in electrical engineering. His M.S. thesis addressed digital restoration of old recordings. From 1999 to 2000, he worked on research and development of a DSP system for analysis classification of sonar signals as part of a cooperation project between the Signal Processing Laboratory (COPPE-UFRJ) and the Brazilian Navy Research Center (IPqM). Since 2000, he has been with the Laboratory of Acoustics and Audio Signal Processing at Helsinki University of Technology, where he is currently pursuing postgraduate studies. He is a grant holder from CNPq, a Brazilian governmental council for funding research in science and technology. His research interests include among others digital audio restoration, computational auditory scene analysis, and sound synthesis. Esquef is an associate member of the IEEE and member of the Audio Engineering Society.



Matti Karjalainen was born in Hankasalmi, Finland, in 1946. He received the M.S. and the Dr.Tech. degrees in electrical engineering from the Tampere University of Technology, in 1970 and 1978, respectively. From 1980, he has been Professor in acoustics and audio signal processing at the Helsinki University of Technology in the faculty of Electrical Engineering. In audio technology, his interest is in audio signal processing, such as DSP for sound reproduction, perceptually based signal processing, as well as music DSP and sound synthesis. In addition to audio DSP, his research activities cover speech synthesis, analysis, and recognition, perceptual auditory modeling and spatial hearing, DSP hardware, software, and programming environments, as well as various branches of acoustics, including musical acoustics and modeling of musical instruments. He has written more than 300 scientific and engineering articles or papers and contributed to organizing several conferences and workshops. Professor Karjalainen is an AES (Audio Engineering Society) Fellow and member in IEEE (Institute of Electrical and Electronics Engineers), ASA (Acoustical Society of America), EAA (European Acoustics Association), ISCA (International Speech Communication Association), and several Finnish scientific and engineering societies.



Vesa Välimäki was born in Kuorevesi, Finland, in 1968. He received his M.S. in technology, Licentiate of Science (Lic.S.) in Technology, and Doctor of Science (D.S.) in Technology degrees in electrical engineering from Helsinki University of Technology (HUT), Espoo, Finland, in 1992, 1994, and 1995, respectively. Dr. Välimäki worked at the HUT Laboratory of Acoustics and Audio Signal Processing from 1990 until 2001.



In 1996, he was a Postdoctoral Research Fellow with the University of Westminster, London, UK. During the academic year 2001–2002, he was Professor of signal processing at Pori School of Technology and Economics, Tampere University of Technology (TUT), Pori, Finland. In August 2002, he returned to HUT where he is currently Professor of audio signal processing. In 2003, he was appointed Docent in signal processing at Pori School of Technology and Economics, TUT. His research interests are in the application of digital signal processing to audio and music. He has published more than 120 papers in international journals and conferences. He holds 2 patents. Dr. Välimäki is a senior member of the IEEE Signal Processing Society and a member of the Audio Engineering Society and the International Computer Music Association.