

# Frequency-Zooming ARMA Modeling of Resonant and Reverberant Systems\*

MATTI KARJALAINEN,<sup>1</sup> AES Fellow, PAULO A. A. ESQUEF,<sup>1</sup> AES Member, POJU ANTSALO,<sup>1</sup>  
AKI MÄKIVIRTA,<sup>2</sup> AES Member, AND VESA VÄLIMÄKI,<sup>3,1</sup> AES Member

<sup>1</sup>Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, FIN-02015, Espoo, Finland

<sup>2</sup>Genelec Oy, FIN-74100 Iisalmi, Finland

<sup>3</sup>Tampere University of Technology, Pori School of Technology and Economics, FIN-28101 Pori, Finland

Discrete-time analysis and modeling of reverberant and resonating systems has many applications in audio and acoustics. The methodology of ARMA modeling by pole-zero filters for measured impulse responses was investigated. In addition to an overview of the standard AR and ARMA techniques, a spectral zooming technique is proposed, which is useful for resolving very closely positioned modes and high-density modal clusters. Application cases related to the analysis and modeling of room responses, loudspeaker-room equalization, and the estimation of parameters for musical instrument modeling are studied.

## 0 INTRODUCTION

Parametric analysis and modeling is an increasingly common task in acoustics and audio. In this paper we focus on audio-related problems where a target system response can be measured and the task is to model it for computational simulation or synthesis, or to derive an inverse model for equalization. Representative examples of the first group are room response modeling, including artificial reverberation design, or just estimation of eigenmodes at low frequencies in room acoustical studies, and modeling of musical instruments. The second category, inverse modeling, is common in audio, where equalization of nonideal response properties is a frequent task in high-quality sound reproduction as well as prevention of acoustic feedback in sound-reinforcement systems.

The behavior of acoustic or audio systems at low frequencies can often be modeled analytically and parameterized, at least with lumped models. Examples of such cases are low-frequency modal behavior in a rectangular room [1], musical instrument body [2], or loudspeaker enclosure [3]. For irregular structures or higher frequencies it is much more difficult or impossible to find analytical or numerical models that are useful in practice. In such cases it is still possible to measure system responses and to apply signal modeling techniques to analyze, sim-

ulate, or synthesize in real time a given response.

Particular interest of this study is focused on the resonant and reverberating characteristics of complex acoustic systems. The modal behavior, that is, the decomposition of eigenmodes assuming a linear and time-invariant (LTI) system, can be extremely complicated. In one-dimensional resonators, such as strings and tubes in musical instruments, the density of eigenmodes is not necessarily high, but these modes can exhibit complicated details, such as two-stage decay and strong beating in a decaying envelope. In two- and three-dimensional resonators, such as membranes, plates, and enclosed spaces, the modal density increases toward higher frequencies, resulting in dense modal patterns and reverberation-type behavior when neighboring eigenmodes overlap remarkably. Also the temporal behavior becomes complex, especially in rooms where direct sound and early reflections are followed by an increasing density of reflections resulting in late reverberation. The modeling of measured responses may benefit from a time-frequency viewpoint where the properties of auditory perception are taken into account.

A rich literature exists on signal modeling of LTI systems [4]–[7] in many branches of engineering, systems sciences, and applications. Software tools for modeling are available, for example, in MATLAB.<sup>4</sup> Here we assume that target systems and desired models, in addition to

\* Manuscript received 2002 June 21. An earlier version of this paper was presented at the 112th Convention of the Audio Engineering Society, Munich, Germany, 2002 May 10–13.

<sup>4</sup>In this study we utilize in particular the functions found in the Signal Processing Toolbox [8]. Another MATLAB toolbox of interest is the System Identification Toolbox [9].

being linear and time invariant, are also stable and causal. We also assume that the measured system responses are not heavily noise contaminated so that estimating the system model parameters is practical for the applications at hand. Then a measured impulse response  $h(n)$  can be approximated by a rational expression in the  $z$ -transform domain,

$$H(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_N z^{-N}}{1 + a_1 z^{-1} + \dots + a_P z^{-P}} = \frac{\sum_{k=0}^N b_k z^{-k}}{1 + \sum_{k=1}^P a_k z^{-k}} \quad (1)$$

which makes it possible to simulate or synthesize the target system efficiently by various digital filter implementations [5], [10], [11] of the estimated transfer function.

In this paper we only deal with discrete-time representations for digital signal processing. Thus the easiest way to “model” a measured response  $h(n)$  or its truncated or windowed version is to take it directly as a finite impulse response (FIR) filter  $H(z) = \sum_{k=0}^N h(k)z^{-k}$ . For complex systems, the length of the finite impulse response required for suitable representation may be too long, preventing real-time implementations. On the other hand, shortening the filter length reduces the capabilities of identifying the inherent resonant properties of the system under study. Filters for infinite impulse response (IIR) can come in two forms: 1) all-pole models where the numerator of Eq. (1) is reduced to a single gain coefficient  $b_0$ , or 2) pole-zero models with both the numerator and the denominator being nontrivial polynomials of  $z$ .

In systems science and engineering, such as in control theory for estimation and identification tasks, the terms autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) have been used for modeling processes similar to all-pole, FIR, and pole-zero filter behavior, respectively. For the sake of convenient use of the abbreviations AR and ARMA, as well as to draw attention to the rich knowledge from various fields other than digital audio signal processing, we apply the terms MA, AR, and ARMA here when referring to specific types of models.

Our aim does not stop at obtaining a useful approximation of a measured target system by a transfer function of the type of Eq. (1). We are interested in decomposing it into a parametric description of its constituent components, particularly the complex-conjugate pole pairs, that is, the complex-valued roots of the denominator polynomial, which represent the eigenmodes of the system and result in the resonant and reverberant behavior. In theory pole pairs are common to all responses in a distributed system such as a room [12] or an instrument body, whereas zeros (roots of the numerator  $\sum_{k=0}^N b_k z^{-k}$ ) are essentially position dependent. We are interested in an accurate estimation of the modal parameters, such as pole angle and radius or, equivalently, mode frequency and decay time constant.

In a recent paper [13] we studied this problem of modal parameter estimation using traditional time–frequency analysis techniques by first trying to isolate potential mode frequencies and then estimating the modal decay

rate from a spectrogram, such as short-time Fourier analysis or cumulative decay spectrum [14]. Decay rate estimation was also applied to wide-band signals, for example, to a robust estimation of the reverberation time  $T_{60}$ . A problem with such methods is to model overlapping modes that result in nonexponential decay in any reasonable frequency span. AR and ARMA approaches try to model the target response globally by minimizing a given modeling error criterion, typically a least-squares error. Thus the interactions of overlapping modes are taken into account simultaneously and systematically.

One problem of straightforward optimization is the inflexibility of global optimization criteria, for example, to take into account varying properties of different frequency ranges. Also selecting proper values for the order parameters  $N$  and  $P$  of Eq. (1) is not easy. A practical problem is that solving the (complex-valued) roots of a high-order polynomial is an ill-posed numerical task. To avoid problems with high-order models we propose a method where a part of the given audio frequency range is modeled at a time to obtain an accurate description of the modes within this frequency span. Frequency-zooming ARMA (FZ-ARMA) modeling is shown to be a powerful way to decompose highly complex resonant responses into modal representations, and related IIR filter implementations can be used for the simulation and synthesis of such systems. Our research is influenced by earlier studies on selective linear prediction [15], multiband modeling of musical signals [16], room responses [17], loudspeaker responses [18], and other high-resolution system modeling techniques.

This paper is structured as follows. Sections 1 and 2 present an overview of the AR and ARMA modeling methods and techniques. Examples are given to illustrate the modeling ability and limitations of these basic approaches. Section 3 introduces the FZ-ARMA method, which is able to analyze high-order systems with overlapping modes and dense modal distributions. The effects of nonidealities, including noise and non-LTI behavior, are discussed. Three cases of audio applications that use AR, ARMA, and FZ-ARMA methods are described in Section 4, including modeling of measured room responses, inverse modeling and equalization of loudspeaker–room responses, and modeling as well as sound synthesis of musical instruments. The final section contains a discussion and concluding remarks.

## 1 AR MODELING

The impulse response of a resonant system shows one or more exponentially decaying sinusoids. Each such “mode” can be inherently modeled by a complex-conjugate pole pair, which suggests AR modeling with the corresponding IIR filters. There is a long tradition for finding the least-squares optimal fit of such models to measured LTI system responses, either to a given impulse response or to input–output signal pairs. Here we briefly refer to the theory of linear prediction (LP), which in particular has found application as a powerful spectral modeling technique in speech processing [19], [20].

### 1.1 Linear Prediction

In linear prediction a signal sample  $x(n)$  is assumed to be predictable as a linear combination of previous samples  $\hat{x}(n) = \sum_{i=1}^P a_i x(n-i)$ . When the least-squares prediction error between  $x(n)$  and  $\hat{x}(n)$  is minimized, the (auto)correlation coefficients

$$r_x(k) = \sum_{i=0}^{P-1} x(i) x(i+k) \tag{2}$$

play a central role. The most frequently used version of linear prediction analysis is the autocorrelation method, where the optimal values of the model parameters  $a_i$  are solved from a linear matrix equation (normal equations),

$$\begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{P-1} \\ r_1 & r_0 & r_1 & \cdots & r_{P-2} \\ r_2 & r_1 & r_0 & \cdots & r_{P-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{P-1} & r_{P-2} & r_{P-3} & \cdots & r_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_P \end{bmatrix} \tag{3}$$

Parameters  $r_k$  are the autocorrelation coefficients  $r_x(k)$  from Eq. (2) for a signal frame under study and  $P$  is the order of linear prediction analysis (order of the all-pole model filter). The coefficients  $a_i$  are the estimated polynomial coefficients in the denominator of Eq. (1), that is,  $1 + \sum_{i=1}^P a_i z^{-i}$ , the numerator being only a gain coefficient.<sup>5</sup>

### 1.2 Limitations of AR Modeling

A problem with the AR modeling of real-world systems, in the formulation described, is that the method is not able to do time-domain fitting to a given target response unless the target process is strictly of the AR (all-

pole) type. Fig. 1 illustrates this clearly in a simple case of one idealized mode. For the damped sinusoid in Fig. 1(a) the model response in Fig. 1(e) is a very accurate replica of the target with model order 2 (one complex-conjugate pole pair). If the initial phase is changed 90 degrees to a damped cosine function, the second-order AR model response (solid line) in Fig. 1(f) deviates severely from the given target in Fig. 1(b).

Further insight into the behavior of AR modeling is gained by noticing that the model is based entirely on the autocorrelation coefficients, which in the frequency domain corresponds to the power spectrum. This means pure spectral modeling, whereby the spectrum of the cosine case in Fig. 1(d) clearly deviates from the spectrum of the sine case response in Fig. 1(c). Irregular onsets are common in acoustic system responses, thus indicating that simple AR modeling will have difficulties, and that more powerful methods are needed for accurate temporal modeling.

A somewhat better match to the decaying cosine tail mentioned can be achieved by the covariance method of AR analysis,<sup>6</sup> but the phase matching problem remains and requires ARMA modeling.

## 2 ARMA MODELING

ARMA modeling, which yields a pole-zero filter, has more modeling power than the AR method. It is, however,

<sup>5</sup>Linear predictive analysis is computed in MATLAB by the function `lpc`, which first solves the autocorrelation coefficients  $r_k$  and then inverts the Toeplitz-type correlation matrix to yield predictor coefficients  $a_i$  through the `levinson` recursion function [8].

<sup>6</sup>Covariance method function `ar`, option `ls`, is found in the System Identification Toolbox of MATLAB.

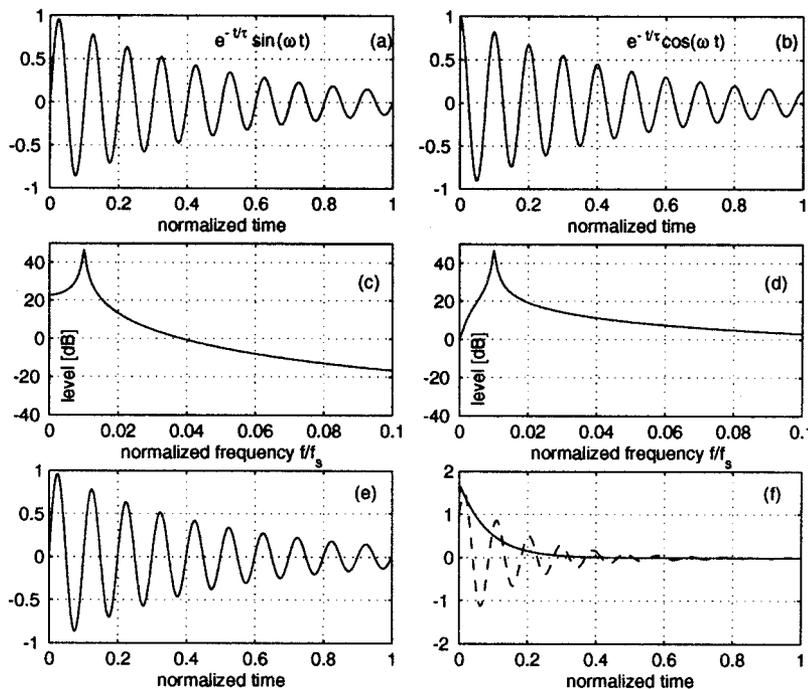


Fig. 1. AR modeling of single-mode decay with different initial phase. (a) Damped sinusoid target response. (b) Damped cosine target response. (c) Magnitude spectrum of sine response. (d) Magnitude spectrum of cosine response. (e) AR model response for sinusoidal case of order 2. (f) AR model response for cosine case of order 2 (—) and order 20 (---).

more difficult because no closed-form solutions are available, thus requiring nonlinear optimization. ARMA estimation algorithms are iterative, starting from a good AR model and then iterating toward optimal parameter values. As with any nonlinear optimization method, a lack of convergence or trapping to a local optimum may occur, and plain computational problems due to insufficient numerical precision are found.

Two methods are applied in the following: Prony's method and the Steiglitz–McBride method.<sup>7</sup> A brief discussion of the problem of model order selection is followed by motivation for the need for improved frequency resolution, before the frequency-zooming ARMA technique is introduced.

### 2.1 Prony's Method

Prony's method [21, pp. 226–228] is a stepwise algorithm that fits  $N + 1$  first samples of a given response exactly, while  $P$  poles of the denominator in Eq. (1) take care of tail decay fitting. Because the AR part estimation is of the covariance type, the resulting filter can become unstable, even in cases where the target system to be modeled is stable.

### 2.2 Steiglitz–McBride Iteration

The Steiglitz–McBride method [7, pp. 174–177] is an algorithm with iterative prefiltering for the least-squares fit of an ARMA model of Eq. (1) to a given impulse response or a given input–output pair (system identification problem). An initial estimate for the denominator can be obtained, for example, by Prony's method.

As with Prony's method, the resulting filter from the

Steiglitz–McBride iteration can be unstable, especially with high-order filters, even for stable target systems. Often the model response starts in a good match with the given time-domain response (since this is least-squares fitting), but after some time it starts to explode due to a pole or poles outside the unit circle.

### 2.3 Model-Order Selection

Both AR and ARMA models need a careful selection of the filter orders  $P$  and  $N$  ( $N = 0$  for AR models).<sup>8</sup> There is no general and automatic way to select optimal filter orders. Rather they can be searched for by various rules to obtain a good enough match to a given target response [22], or the orders can be approximated using a priori information about the target system.

An illustrative example on how the order of AR modeling (linear prediction) affects the estimated modal frequencies (pole angles) is shown in Fig. 2. The magnitude spectrum of a measured room response is plotted in comparison with a related map that shows the pole frequencies for AR model orders up to 100. For lowest orders only the most prominent spectral peaks become roughly approximated, and for increasing model orders these poles split into new pole pairs and groups of poles.

In this study we are interested in modeling resonant and reverberant systems by methods where the poles and the related parameters, angles and radii, can be resolved explicitly. This is needed in applications such as those discussed in Section 4. Selection of the model order is then more demanding than in cases where finding the numerator and denominator polynomials is enough. If the model order is too low, not all modes are represented properly by complex-conjugate pole pairs, or the radii of the poles found remain underestimated. If a model order is too high, single modes often become overmodeled, that is, more than one pole pair will become allocated per mode.

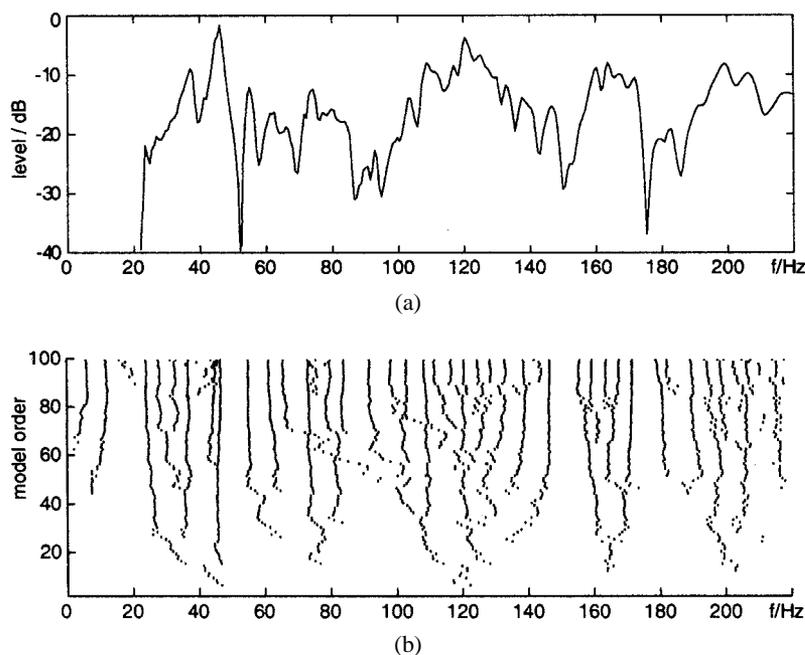


Fig. 2. (a) Magnitude spectrum of measured room response. (b) Frequencies corresponding to pole angles obtained from linear prediction of varying order.

## 2.4 Limitations of ARMA Modeling

While powerful in simple cases for low-order models, the methods described in the preceding exhibit difficulties with high-order modeling of complex target systems. Often these problems originate from limited computational precision. In Prony's method and the Steiglitz–McBride iteration potential instability is often a problem. Although poles outside the unit circle can be mirrored inside the unit circle, yielding an equivalent-magnitude spectrum, the temporal structure of the impulse response is changed.

Linear prediction (autocorrelation method) may yield stable and accurate results with model orders of hundreds or thousands, in particular when the poles are not very close to the unit circle or to each other. However, if the poles (and zeros) must be solved explicitly, numerical problems arise, as mentioned. Frequency-selective AR and ARMA modeling can solve some of these problems.

## 3 FREQUENCY-ZOOMING ARMA (FZ-ARMA)

Problems with resolving very closely positioned modes and mode groups were the reason for experimenting with methods that have better control over frequency resolution. Several ideas are available for improvement, including frequency warping [23] and frequency-selective modeling such as selective linear prediction [15], multiband AR/ARMA techniques [16], and many other high-resolution signal modeling methods.

Frequency warping is a convenient technique when either the lowest or the highest frequencies require enhanced frequency resolution. This approach can be extended to Kautz filters, which exhibit interesting properties of generalized frequency resolution control [24]. These methods have, however, been left out of the scope of this study.

Frequency-selective modeling has been applied, for example, in the linear prediction of speech. In a simple case a target response can be low-pass filtered and decimated in order to model the low-frequency part of the response. A range of higher frequencies can be modulated down and decimated prior to similar modeling. Actually any subband of a given frequency range can be modeled this way, and finally the resulting parameters (poles and zeros) can be mapped back to the original sample rate domain. This is called here modeling by frequency zooming. It resembles the multiband/subband techniques used in [16], [18], [25], [26].

### 3.1 Formulation of Frequency Zooming

The FZ-ARMA (or FZ-AR) analysis starts by modulating (heterodyning) the desired frequency band of impulse response  $h(n)$  down to the neighborhood of zero frequency [27]–[29] by

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

where  $\Omega_m = 2\pi f_m/f_s$ ,  $f_m$  being the modulation frequency and  $f_s$  the sample rate. In the  $z$  domain this can be inter-

preted as clockwise rotation of poles  $z_i$  by angle  $\Omega_m$ , that is,

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

but retaining the pole radius. The next step to increase the frequency resolution is to limit the frequency range by decimating, that is, low-pass filtering and down-sampling the rotated response by the zooming factor  $K_{\text{zoom}}$  to obtain a new sampling rate  $f_{s, \text{zoom}} = f_s/K_{\text{zoom}}$ . This implies mapping to a new  $z$  domain, where poles are scaled by the rule

$$z_{i, \text{zoom}} = z_i^{K_{\text{zoom}}} \quad (6)$$

Together the mappings [Eqs. (5) and (6)] yield the new poles,<sup>9</sup>

$$\hat{z}_{i, \text{zoom}} = |z_i|^{K_{\text{zoom}}} \exp\left\{j\left[\arg(z_i) - \Omega_m\right]K_{\text{zoom}}\right\} \quad (7)$$

Now it is possible to apply any AR or ARMA modeling to the modulated and decimated response. Notice that this new signal is complex-valued due to the one-sided modulation operation.

The advantage gained by frequency zooming is that in the zoomed subband the order of (ARMA) analysis can be reduced by increasing the zooming factor  $K_{\text{zoom}}$ , and consequently, the solution of poles and zeros as roots of the denominator and numerator polynomials of the model function, Eq. (1), is simplified. In addition this means that a different resolution can be used in each subband, based, for example, on knowledge about the modal complexity of a subband.

After the poles have been solved within a zoomed subband, they can be remapped to the full sample rate by inverse scaling the pole radii as well as rotating them counterclockwise,

$$\hat{z}_i = \hat{z}_{i, \text{zoom}}^{1/K_{\text{zoom}}} e^{j\Omega_m} \quad (8)$$

Because of the one-sided down modulation used in Eq. (4), each pole  $\hat{z}_i$  must be used as a complex-conjugate pair in order to obtain real-valued filters.

Finally there are two alternatives of complete model construction. A full audio range model can be realized as a decimated filter-bank implementation or the subband models can be combined into a full-rate filter. The details of subband realization may follow any of the known multirate techniques [30]. For a full-rate case, the final step is to combine the poles and zeros obtained from different subbands. This is a nontrivial task, not discussed in detail here. It is advantageous to pick poles only within the central parts of overlapping subbands to avoid poles due to the boundaries of subbands needed for band limitation. The processing of the zeros is discussed briefly in the examples of this section.

In the investigations of FZ-ARMA that follow, the frequency-zooming method used for solving ARMA coef-

<sup>9</sup>Note that Eqs. (5) and (6) merely characterize how the  $z$ -domain properties of a given response are changed through modulation and decimation, but the estimated pole-zero pattern of an FZ-ARMA model will be obtained only in the next step.

ficients is the Steiglitz–McBride iteration. Notice that the filter orders  $N$  and  $P$  refer to real-valued filters, with complex-conjugate pairs constructed from one-sided zeros and poles obtained from the model of the decimated signal. Thus the orders of real-valued filters are twice the numbers of zeros and poles from the procedures described.

### 3.2 Modeling of Higher Order Modal Decays

In this section the performance of the FZ-ARMA analysis is illustrated through synthetic signals. In particular we are interested in investigating the modeling capability when dealing with signals exhibiting beating or two-stage decay in their envelopes. Simple signals featuring these characteristics can be obtained by

$$s(n) = \sum_{k=1}^M a_k e^{-n/(f_s \tau_k)} \sin \left( 2\pi n \frac{f_k}{f_s} + \theta_k \right) \quad (9)$$

where  $M$  is the number of modal frequencies present in  $s(n)$ ,  $\tau_k$  are decay time constants,  $f_k$  are modal frequencies,  $f_s$  is the sampling frequency, and  $\theta_k$  are the initial phases of the modes.

Let us start with case A, in which the amplitude envelope of a signal consisting of two modes shows beating. The parameters used to generate the signal as well as those adopted in FZ-ARMA modeling are given in Table 1. The

Table 1. Parameters in simulation case A.

Synthetic Signal							
$f_1$ (Hz)	$f_2$ (Hz)	$a_1$	$a_2$	$\tau_1$ (s)	$\tau_2$ (s)	$\theta_1$	$\theta_2$
100	115	0.5	0.5	0.07	0.07	0	0
FZ-ARMA							
$K_{\text{zoom}}$	$f_m$ (Hz)	$N$	$P$				
228	107.5	4	4				

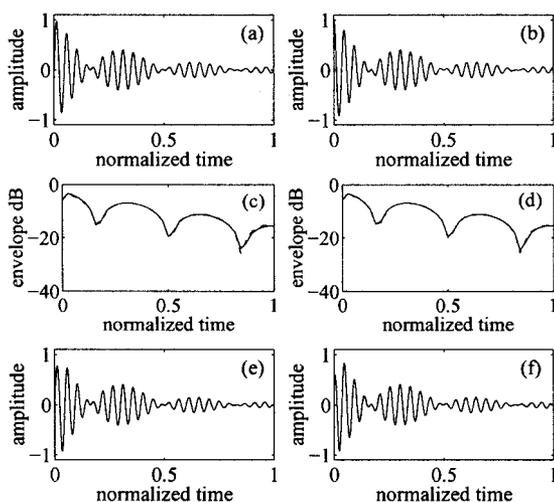


Fig. 3. Case A, FZ-ARMA modeling of amplitude beating due to two modes very near in frequency. (a) Synthetic signal generated according to Eq. (9) with parameters given in Table 1. (c) Original (---) and modeled (—) amplitude envelopes (curves overlap almost perfectly). (e) Resynthesized signal based on estimated model. (b), (d), (f) Another signal generated with parameters given in Table 1, but with phases replaced by  $\theta_1 = \theta_2 = \pi/2$ .

target responses in sine and cosine phases, their FZ-ARMA envelopes, and their resynthesized versions are shown in Fig. 3. The envelopes are obtained from the complex decimated signals by taking the absolute values.

Each resynthesized response is computed by mapping back the pole–zero model impulse response from the decimated domain to the full sample rate. This is done by interpolation and low-pass filtering (using the MATLAB function `resample`), one-sided demodulation, taking the real part, and multiplying by 2 (to compensate the missing complex-conjugate part), that is, by applying the inverse operations of the previous section. For an efficient realization of FZ-ARMA models it is advantageous to implement them as decimated subband filters, but this topic is beyond the scope of this paper.

In the simulation results of case A in Fig. 3, an ARMA (4, 4) model suffices to represent properly the envelope decays in Fig. 3(c) and (d), while the initial phase characteristics of the resynthesized signal in Fig. 3(f) deviate from those in Fig. 3(b). Note that it is almost impossible to distinguish between the dashed and solid lines in Fig. 3(c) and (d).

In case B we verify the FZ-ARMA modeling of a two-mode response for which the amplitude envelope exhibits a two-stage decay. The parameters used to generate this signal, as well as those of FZ-ARMA modeling, are summarized in Table 2, and the results of the modeling are shown in Fig. 4. The slower decaying mode is modeled properly although its initial level is 10 dB below the stronger one. This capability of two-stage decay analysis can work down to  $-30$  dB in a clean synthetic case.

### 3.3 Modeling of Noisy Responses

In simulation case C, in order to verify the FZ-ARMA modeling when dealing with noisy signals, we contaminate the impulse responses shown in Fig. 3(a) and (b) with zero-mean additive white Gaussian noise. In this example the variance of the noise is chosen to produce a signal-to-noise ratio (SNR) of  $-5$  dB in the beginning of the signal. Of course, the local SNR decreases toward the end of the signal.

The results are displayed in Fig. 5, which follows the same structure as the previous figures. Looking at Fig. 5(c) and (d) it can be seen that the envelopes of the modeled signals (solid lines) differ substantially from those of the noisy signals (dashed lines). Moreover, the resynthesized signals based on the computed models, shown in Fig. 5(e) and (f), are free of visible noise and follow closely their corresponding clean versions, which are depicted in Fig. 3(a) and (b).

Table 2. Parameters in simulation case B.

Synthetic Signal							
$f_1$ (Hz)	$f_2$ (Hz)	$a_1$	$a_2$	$\tau_1$ (s)	$\tau_2$ (s)	$\theta_1$	$\theta_2$
100	100	0.2	0.8	0.3	0.02	0	0
FZ-ARMA							
$K_{\text{zoom}}$	$f_m$ (Hz)	$N$	$P$				
245	100.0	6	6				

The highly successful result of reducing the additive noise in simulation case C can be understood when considering the frequency zooming to a narrow band around the modal frequencies of interest, whereby the SNR is improved by the zooming ratio, that is, by  $10 \log_{10} 228 = 23.6$  dB in this case. Low-order ARMA (4, 4) modeling further reduces the influence of noise due to a good correlation with the modal signals only.

### 3.4 Envelope Modeling of Non-LTI Systems

A primary assumption when applying FZ-ARMA or any LTI system modeling is that the frequencies of the modes do not change within the duration of the analyzed segment. Even if this requirement cannot be satisfied, for example, in strongly plucked string instrument tones having initial pitch shifting [31], the envelope behavior of the target signal can still be modeled. A straightforward way, if the frequency trajectory of the pitch shift is known, is to resample the signal so that the shift is eliminated. Another way is to apply FZ-ARMA modeling but adopt higher orders for the numerator and denominator so that this can capture the effect of the frequency shift.

Alternatively one can compute an ARMA model for the envelope of a modulated and decimated signal (FZ-ENV-ARMA). In that way the envelope behavior can be approximated with a lower model order. Simulation case D compares standard FZ-ARMA modeling with FZ-ENV-ARMA, as shown in Fig. 6. The test signal plotted in Fig. 6(a) is a variant of the one plotted in Fig. 3, but now the initial values of the mode frequencies start 50 Hz above the values indicated in Table 1, and then they shift exponentially with a time constant of 100 ms to the nominal values. The subplots in the left column show original and modeled envelopes for different FZ-ENV-ARMA model orders. The subplots in the right column do the same, but using FZ-ARMA models.

To resynthesize a changing pitch signal based on the FZ-

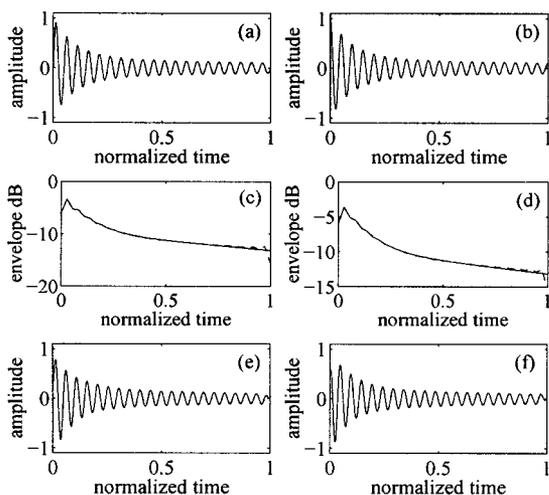


Fig. 4. Case B, FZ-ARMA modeling of two-stage decay due to two modes with equal frequencies. (a) Synthetic signal generated according to Eq. (9) with parameters given in Table 2. (c) Original (---) and modeled (—) amplitude envelopes. (e) Resynthesized signal based on estimated model. (b), (d), (f) Another signal generated with parameters given in Table 2, but with phases replaced by  $\theta_1 = \theta_2 = \pi/2$ .

ENV-ARMA computed model, it is necessary to estimate its pitch behavior. Then, after obtaining a model for the amplitude envelope, a frequency modulation corresponding to the original frequency shift should be employed during synthesis. For direct FZ-ARMA modeling this is not needed as long as the estimation is capable of capturing the given behavior of the shifting modal frequencies.

It can be verified from Fig. 6 that, in contrast to what happens with FZ-ARMA modeling, increasing the model

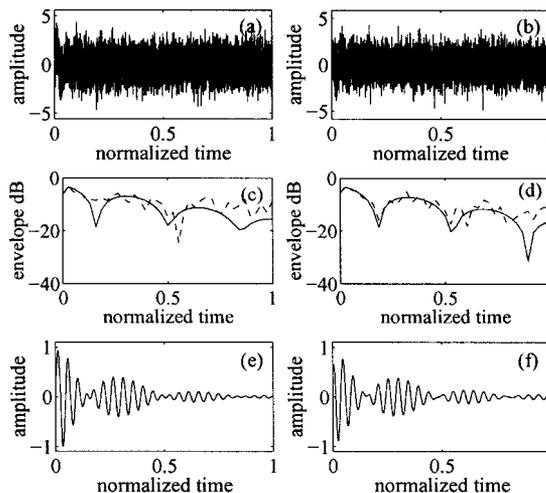


Fig. 5. Case C, FZ-ARMA modeling of amplitude beating of two modes in noise. (a) Synthetic signal generated according to Eq. (9) with parameters given in Table 1, immersed in white noise. (c) Original (---) and modeled (—) amplitude envelopes. (e) Resynthesized signal based on estimated model. (b), (d), (f) Another signal generated with parameters given in Table 1, but with phases replaced by  $\theta_1 = \theta_2 = \pi/2$ .

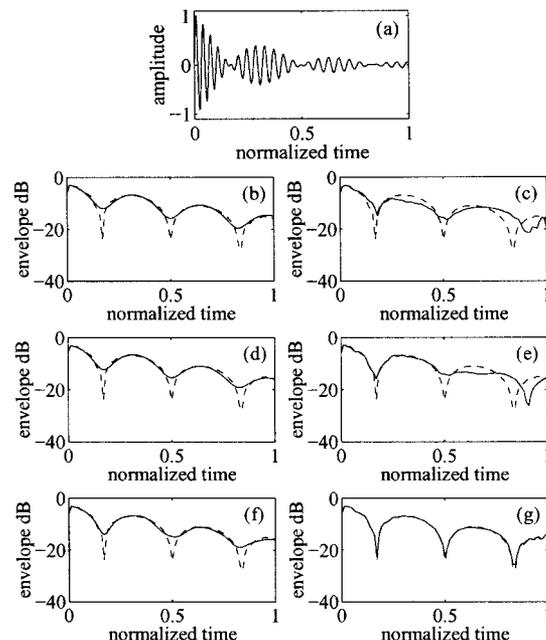


Fig. 6. Case D, FZ-ARMA modeling of amplitude beating with pitch shift. (a) Synthetic signal generated according to Eq. (9) with parameters given in Table 1, but with a pitch shift. (b), (d), (f) Original (---) and modeled (—) amplitude envelopes for FZ-ENV-ARMA models. (b) ARMA (4, 4). (d) ARMA (8, 8). (f) ARMA (12, 12). (c), (e), (g) Same cases but for FZ-ARMA modeling, with corresponding model orders.

order in FZ-ENV-ARMA does not help to improve the model fit substantially since the inherent phase relations of the original signal have been lost in the computation of the envelope that is used as a target. Nevertheless, for low-order modeling, FZ-ENV-ARMA yields a better envelope fit than equal-order FZ-ARMA modeling.

If the response of a target system is of highly complex mode density, such as a room response at medium to high frequencies, a detailed modal description may not be feasible or desired. In such cases the envelope behavior can be represented simply by fitting a lower order model to the decaying envelope in a desired frequency range by FZ-ENV-ARMA techniques. This can be useful in decay time estimation.<sup>10</sup> Simulation case E in Fig. 7 depicts the decay envelope of an example room response for the octave band of 1–2 kHz and a related envelope curve fitting by low-order FZ-ARMA modeling.

Another form of non-LTI behavior is nonlinearity. A small degree of nonlinearity in a system can be accepted, and even a quite severe deviation from linearity can be tolerated if we accept the fact that the parameters are then signal dependent, such as dependent on the level of a signal.

### 3.5 Selection of Zooming Parameters

The choice of the FZ-ARMA parameters  $\Omega_m$  and  $K_{zoom}$ , and of the ARMA orders  $N$  and  $P$  depends on several factors. Considering first the zoom factor, it can be said that the larger  $K_{zoom}$  is, the higher the frequency resolution. This favors cases in which the modes are densely distributed in frequency. On the other hand, high values for  $K_{zoom}$  imply a more demanding signal decimation procedure and fewer samples available for modeling in the decimated signal.

A natural choice when there are relatively isolated modes or mode groups is to select the frequency range of focusing to cover such a group and its vicinity until the neighboring modes or groups start to have an influence. It is recommended to choose the range of focus such that resonance peaks are not placed at the edges of the focused subband. As a rule of thumb, a suitable choice is to set  $f_m$  to the middle of the subband of interest.<sup>11</sup> This frequency is then mapped to the zero frequency of the decimated frequency range.

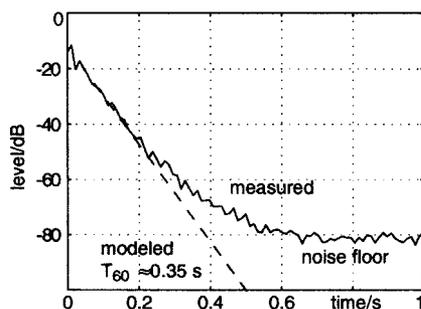


Fig. 7. Case E, estimation of reverberant response decay rate and  $T_{60}$  by modeling decay envelope using low-order FZ-ENV-ARMA model. Measured response was bandpass filtered (1–2 kHz), absolute value of envelope taken and decimated by 500, and modeled by filter orders  $N = 6$ ,  $P = 6$ . The largest positive (real) pole corresponding to the main decay component was identified for slope estimation.

The order of an FZ-ARMA model will depend on the number of modes associated with each resonance group. Experiments on two-mode resonances reveal that adopting an FZ-ARMA (4, 4) model in general yields satisfactory results for such cases. Better modeling accuracy can be achieved by increasing the order, although the resulting poles may no longer be physically interpretable for a two-mode case. High-order analysis also raises the probability of ending up with an unstable model.

## 4 CASE STUDIES IN AUDIO APPLICATIONS

AR and ARMA modeling have many applications in modern audio signal processing. Linear and time-invariant models can be applied, for example, in room acoustics, sound synthesis, and audio reproduction. Based on the theoretical overview and examples given in the preceding sections, we will study the feasibility of the methods, in particular of the FZ-ARMA technique, in several audio applications.

### 4.1 Modeling of Room Impulse Responses

A challenging application for AR/ARMA modeling is to find compact but perceptually valid approximations for measured (or computed) room impulse responses [12]. This is needed for modal analysis of rooms at low frequencies, artificial reverberation designs, or equalization of loudspeaker–room responses, to name a few. An example of a similar subband modeling technique is found in [17], [33].

In case study F an analysis of the low-frequency modal behavior of a room impulse response is carried out using different AR and FZ-ARMA methods. The room has approximate dimensions of 5.5 by 6.5 by 2.7 m<sup>3</sup>. Fig. 8(a) illustrates the time–frequency behavior (waterfall plot of cumulative spectral decay) for frequencies below 220 Hz, computed from a measured room impulse response. The room shows particularly intense and long-ringing modal decays around 45 Hz.

Straightforward AR modeling of the room impulse response below 220 Hz using linear prediction yields fairly accurate results when the all-pole filter order  $P$  is about 100 or higher. Fig. 8(b) shows the decay plot model response when  $P = 80$ . The original sample rate of 44 100 Hz was reduced by a factor of 110 prior to AR modeling. A comparison with Fig. 8(a) reveals that the decay times of the prominent modes are quite well modeled, but many weaker modes are too short or too damped due to insufficient model order.

Direct ARMA modeling by the Steiglitz–McBride method yields a better time-domain fit with a given denominator order than the corresponding AR model. For example, using a numerator  $N = 30$  and a denominator  $P = 100$  worked fairly well for the previous room response, although in many cases the Steiglitz–McBride algorithm gives

<sup>10</sup>There exist many methods that are better suited for estimating the reverberation time  $RT_{60}$  [13].

<sup>11</sup>In an earlier version of this paper [32] the modulation frequency  $f_m$  was positioned at the lower edge of the subband of interest.

already an unstable result with such moderate filter orders.

FZ-ARMA is a powerful method for accurate modeling of modal behavior in a limited frequency range. Fig. 9 depicts modeling results of the prominent modal region of 33–58 Hz in the response of Fig. 8(a). The region includes three major modes at frequencies of 37, 46, and 55 Hz. Fig. 9 shows the decay envelope of the modal region for the original signal (dashed line) and when applying the Steiglitz–McBride method of different orders (solid lines). Increasing the filter order improves the envelope fit, but finally it may start to model the background noise envelope. The pole–zero plots on the right-hand side indicate that for an order  $P = 6$  the poles correspond to the three modes, whereas for higher orders there are extra poles and it is not easy to associate them with the modes.

At higher frequencies, above the critical frequency (Schroeder frequency) [34] of the room, the modal behavior is diffuse, that is, the modal density is high and the modes overlap in frequency. Full audio range AR and ARMA modeling is difficult, if not impossible, in a single analysis. However, it is possible to apply the FZ-ARMA analysis to any narrow enough frequency band of a rever-

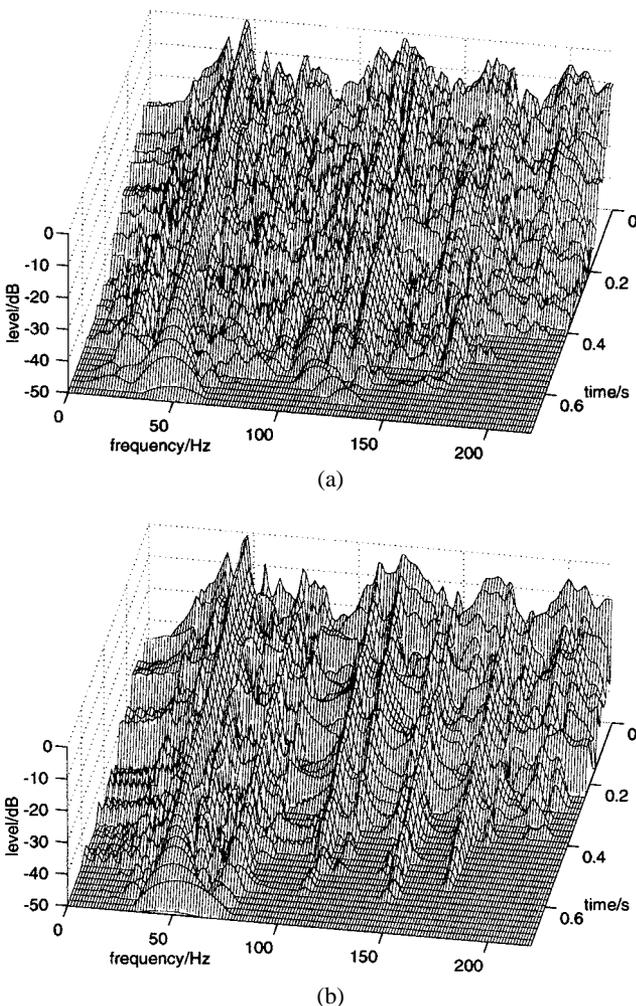


Fig. 8. Case F. (a) Waterfall plot of measured room response below 220 Hz for time span 0.0–1.0 s. (b) Waterfall plot computed from impulse response of AR model ( $P = 80$ ) for room response of (a). Original sampling rate of 44 100 Hz was decimated to 400 Hz before AR modeling. Level is limited to  $-50$  dB.

berant response. Fig. 10 describes a fitting to the room response studied within a critical band at 1 kHz (920–1080 Hz) by different model orders. With the highest model order  $P = 60$ , envelope fitting is good for the first 250 ms and for about a 40-dB dynamic range. Full audio range modeling, based on this subband approach, is discussed in [35].

## 4.2 Loudspeaker–Room Equalization

Equalization of a loudspeaker response or a loudspeaker–room reproduction chain means correcting the system response closer to desired perceptual or technical criteria. MA and ARMA modeling have been reported in the literature in several forms for loudspeaker and in-situ frequency response equalization, in both on-line and off-line formulations [36]–[49], [18], [24].

Equalization of the free-field magnitude response (possibly including the phase response) of a loudspeaker by digital signal processing can be carried out using many known techniques. For highest quality loudspeakers there is hardly any need to improve the free-field response, but the loudspeaker–room combination may benefit greatly from proper equalization.

The combined task of loudspeaker and room equalization is demanding since it is essentially a problem of finding a perceptually optimal time–frequency equalization, instead of simple flattening of the magnitude spectrum or phase linearization. There seems to be quite a common misunderstanding that just flattening the response, at least at low frequencies where it might be technically feasible, would be an ideal solution. A better strategy is to improve the balance of the overall acoustical parameters, particularly of the reverberation time. As discussed in [50], this can be done by controlling the decay times of individual modes at low frequencies, typically below 200 Hz, to match the reverberation time at midfrequencies. This is called modal equalization. It can be followed by a traditional correction of the envelope of the magnitude response. The need for such an active correction of the room acoustics is particularly prominent around 100 Hz, even in spaces designed for listening purposes such as audio monitoring rooms [51].

In [50] we proposed a method for modal equalization. In the present paper we suggest another technique to realize modal equalization, optionally combined with magnitude envelope correction. The general framework of modal equalization has been discussed in detail in the previous paper. A brief description of the procedure follows.

1) Measure the combined loudspeaker plus room impulse response in the listening position of interest. Any modern technique for reliable response measurement can be applied.

2) Analyze the average reverberation time  $RT_{60}$  at midfrequencies, for example, between 500 Hz and 2 kHz.

3) Determine the upper limit of modal decay time as a function of frequency for the low-frequency range, typically below 200 Hz. This value can be allowed to grow slightly toward the lowest frequencies [52], [53], for example, linearly by 0.2 s when the frequency decreases from 300 Hz to 50 Hz.

4) Find the modes that need equalization, that is, those that have a longer decay time than the upper limit defined in step 3. If the magnitude level of a mode is so low that its tail remains below a given level, it does not need modal equalization, even when its decay time is longer than the

upper limit. Estimate the modal parameter values for these modes, in particular the modal frequency and the decay time constant, and compute the angles and radii of the corresponding poles.

5) Design a correction filter for each mode requiring

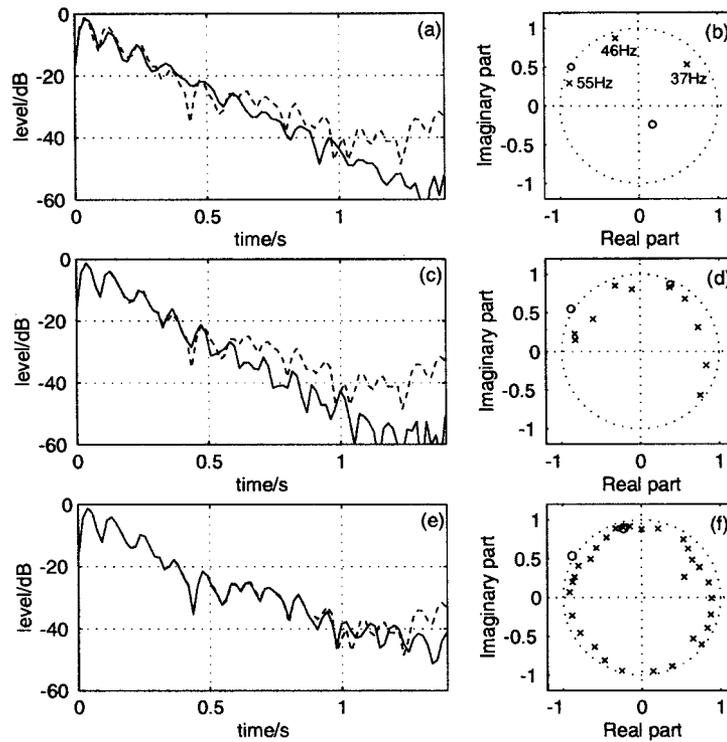


Fig. 9. Case G, envelope match for decay of modal group between 33 and 58 Hz in room response of Fig. 8 with different FZ-ARMA orders. — modeled; - - - original. (a) ARMA (6, 4). (c) ARMA (20, 4). (e) ARMA (60, 6). (b), (d), (f) Poles (x) and zeros (o) in decimated  $z$  domain. (Some zeros fall outside the plotting range.)

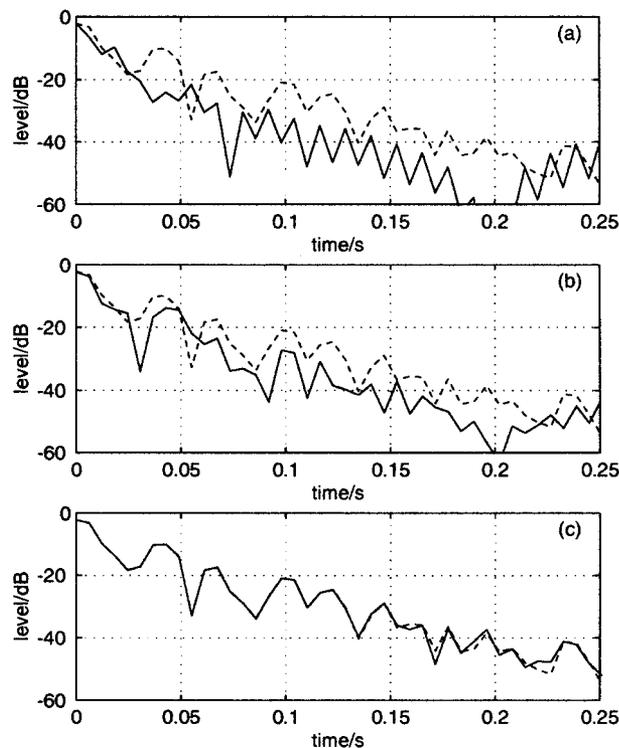


Fig. 10. Case H, fitting to room response within critical band of 920–1080 Hz by FZ-ARMA. Original ( - - - ) and modeled ( — ) responses. (a) ARMA (6, 6). (b) ARMA (20, 6). (c) ARMA (60, 6).

equalization so that the filter shortens the decay time to meet the upper limit criteria specified in step 3. This means canceling the estimated pole pair, which represents a mode with a long decay time, by a zero pair, and replacing it with a new pole pair having the desired decay time. This can be done with an IIR filter [50],

$$H_c(z) = \frac{(1 - r e^{j\theta} z^{-1})(1 - r e^{-j\theta} z^{-1})}{(1 - r_c e^{j\theta_c} z^{-1})(1 - r_c e^{-j\theta_c} z^{-1})} \quad (10)$$

where  $r$  and  $r_c$  are the (complex-conjugate) pole radii of the original decay and the corrected decay, respectively, and  $\theta$  and  $\theta_c$  are the corresponding pole angles.

6) Compute steps 4 and 5 either in a batch mode, that is, in parallel for each mode to be equalized, or iteratively so that the modes are equalized one by one, starting from the most prominent one and returning to step 4, to be applied to the result of the previous equalization. The process is terminated when all remaining modes meet the decay time criteria or when a preset upper limit of correctable modes has been reached.

7) Traditional magnitude equalization can be applied to the result of modal equalization, if needed, by any method or technique appropriate.

In this context we are only interested in step 4 as part of a batch or iterative analysis. All other steps follow the general scheme described in [50], where the mode search and the decay time estimation were based on a time–frequency representation and fitting of a logarithmic decay plus background noise model using nonlinear optimization. While the previously proposed method is found robust for modes that are separated well enough, strongly overlapping or multiple modes with closely similar frequency are an inherent difficulty of that method. Since AR/ARMA models search for a global optimum and do not try to separate modes in the first place, they are potentially a better alternative in such cases.

In the equalization cases that follow, mode finding and parameter estimation are carried out iteratively in the following way.

1) Compute a function that can be used robustly to find the most prominent modes and their frequencies. This can be done in different ways, for example, directly by AR, ARMA, or FZ-ARMA analysis and finding the poles with the largest radii. Because selection of the proper model order can be problematic, we first applied here a separate mode detection function,

$$G(\Omega) = \sqrt{|H(\Omega)| \max(0, D(\arg(H(\Omega))))} \quad (11)$$

where  $H(\Omega)$  is the Fourier transform of the measured response,  $\Omega$  is the normalized angular frequency (angle in the  $z$  plane), and  $D$  is a differentiation operator (in the frequency domain). An example of the  $G(\Omega)$  function is plotted in Fig. 11(d). Positive peaks indicate strong modes that may need decay time equalization. Note that this function combines the information of both magnitude level and decay time (through a phase derivative).

2) Find the highest peak position  $\Omega_p$  of the detection

function  $G(\Omega)$  as the best candidate for modal equalization. Run an AR, ARMA, or FZ-ARMA analysis of predefined order (for simplicity we applied order 70 AR modeling for the low-frequency range up to 220 Hz) on the minimum-phase version of the target response to find poles and select the pole closest to the phase  $e^{j\Omega_p}$  on the unit circle. This pole and its complex conjugate now represent the most prominent mode.

3) If the decay time of the mode is below the upper limit allowed and the value of  $G(\Omega_p)$  is below a threshold determined experimentally, go to finalize the process in step 4. If not, design a second-order modal correction filter of the type of Eq. (10) to change the modal decay time to a desired value below the upper limit. Apply this to the response to be equalized and use the result when going back to iterate from step 1.

4) Finally, collect the correction filters into a cascaded filter, which is now the modal equalizer for the system.

A simulated modal equalization, case I, is illustrated in Fig. 11. A measured loudspeaker impulse response is filtered to add five simulated modes at frequencies of 50, 55, 100, 130, and 180 Hz, with 60-dB modal decay times of 1.4, 0.8, 1.0, 0.8, and 0.7 s, respectively. The waterfall plot of this synthetic response is shown in Fig. 11(a). Fig. 11(b) proves that the effect of the modes can be canceled out almost perfectly, leaving the loudspeaker response only, by moving the pole radii to correspond to a very short decay time (about 60 ms) using the procedure described. In Fig. 11(c) the result of the modal equalization depicted is more appropriate for real room conditions. The decay time of each mode is equalized to 250 ms. The two nearby modes overlap partly at 50 and 55 Hz, and do not cause any difficulties. Hence the modal equalization works almost ideally.

In case J of Fig. 12 the most prominent single mode at 46 Hz in a real room response (Fig. 8) is equalized by shortening the decay time from a value above 1 s to about 300 ms using the algorithm described and limiting the search for modes to only one. In Fig. 12(b) the originally problematic mode decays now clearly faster. Furthermore, the equalized response up to 80 Hz has a much smoother shape since the modal equalization also affects the magnitude spectrum. However, the decay times of some other modes remain long.

Multimode equalization of the same room, case K, is shown in Fig. 13. The procedure described is iterated 100 times, yielding 100 second-order correction filter sections, to shorten the mode decay times. The cumulative decay spectrum of the resulting equalized response is illustrated in Fig. 13(b). The target value for the equalized modal decay time (60 dB) has been 150 ms.

In this case the result is not perfect as in the synthetic case. There is about 10 dB of fast decay in the beginning, as shown by plots of backward integrated energy in Fig. 13(c), and thereafter the decay rate follows the original one. Although the ideal shortening of the decay time is not achieved precisely, it already makes sound reproduction in the room perceptually more balanced in terms of reverberation. Furthermore, the equalization procedure can be improved by careful adjustment of the details.

The final step, namely, smooth envelope equalization of the magnitude response, is not discussed here since many known techniques could be applied to equalize the magnitude response. An interesting choice to be mentioned, however, is to integrate the magnitude equalization stage together with the AR/ARMA modal equalization process.

### 4.3 Modeling of String Instrument Sounds

An appealing application of the FZ-ARMA scheme is the modeling of musical instruments. For instance, in sound synthesis of string instruments by any parametric approach, such as digital waveguide modeling [54], one

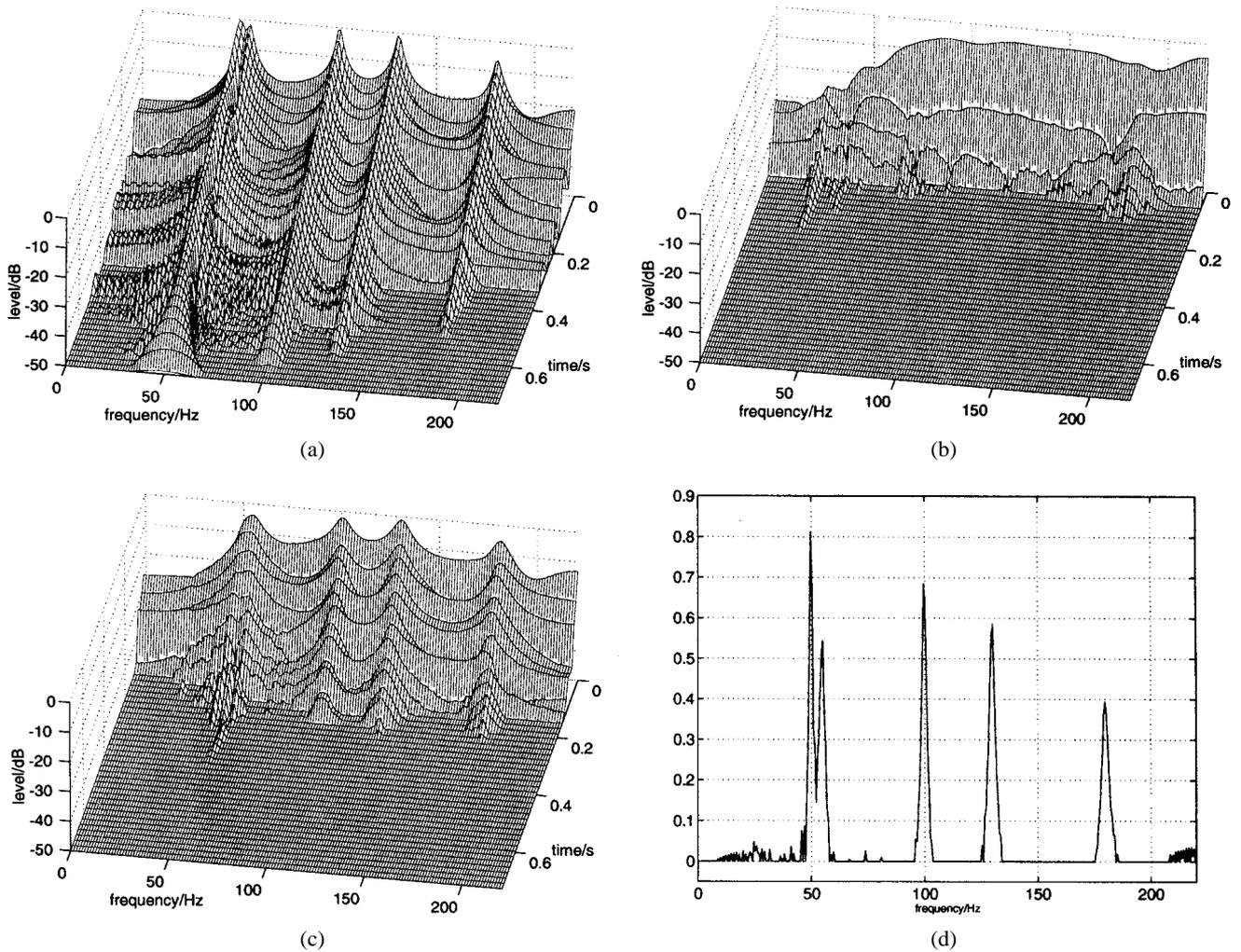


Fig. 11. Case I, waterfall plots for loudspeaker plus synthetic mode equalization. (a) Original response plus five modes at frequencies of 50, 55, 100, 130, and 180 Hz. (b) Modes fully damped. (c) Modal decay time (60 dB) equalized to 250 ms. (d) Mode detection function  $G(\Omega)$  applied to original response. Decay levels in (a)–(c) are limited to  $-50$  dB.

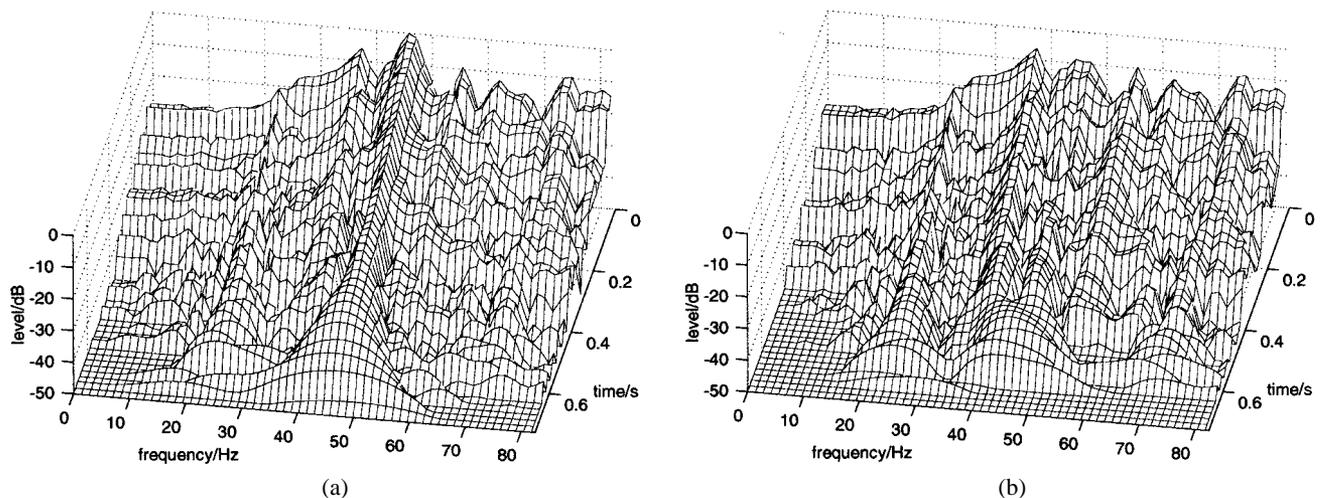


Fig. 12. Case J, equalization of single mode at 46 Hz of room analyzed in Fig. 8. (a) Original response. (b) After mode equalization.

may need to obtain information related to the partial frequencies and their respective decay rates from measured signals [55], [28]. FZ-ARMA modeling turns out to be a suitable tool for such modal analysis purposes.

The experiments with string and bell sounds here have been influenced by techniques published in [56]–[60], [16]. In particular the application of Prony's method [16] is aiming for the same goal. We have found that the Steiglitz–McBride iteration as part of FZ-ARMA yields typically better results than Prony's method.

As a simulated case to study the required orders to model isolated partials we generated a synthetic guitar tone using the dual-polarization model presented in [61]. Each partial has two modes with known parameters, namely, the resonance frequencies and the time constants of the exponentially decaying envelope. In simulations it was found that low-order frequency-focusing models of ARMA (4, 2), that is, with two pole pairs, suffice already to estimate the mode frequencies properly. To match the decay times and forms of the modes accurately, a somewhat higher model order is required. FZ-ARMA (6, 4) worked already well.

When dealing with real instrument tones, such as guitar tones recorded in an anechoic chamber, it is necessary to

estimate the frequencies of the resonance peaks beforehand and select subbands around them in order to proceed with FZ-ARMA modeling.

For harmonic signals with nonvarying pitch, a pitch detector could be used to select both the down-modulation frequencies  $f_m$  and the subband bandwidth  $f_{s,zoom}$ . A more general approach should also work for inharmonic signals, and in this case running a peak-picking algorithm over the magnitude spectrum of the initial part of the tone seems to be more appropriate. In addition, the estimation of the number of resonant modes per partial may be based on a priori information such as the number of vibrating strings per note in a piano tone or the number of polarizations per string.

In an example, case L, FZ-ARMA modeling was performed on a guitar tone. The signal was recorded in an anechoic chamber and corresponds to plucking the fifth string (A2,  $f_0 \approx 110$  Hz). The results of modeling the first 10 harmonics,  $f_i$ ,  $i = 1, \dots, 10$ , of this tone are shown in Fig. 14. The parameters used in FZ-ARMA modeling were  $f_m = f_i$ ,  $K_{zoom} = 600$ ,  $P = 6$ , and  $N = 6$ . The temporal structure of the harmonics is well approximated in this case.

As another example, case M, Fig. 15 plots the original

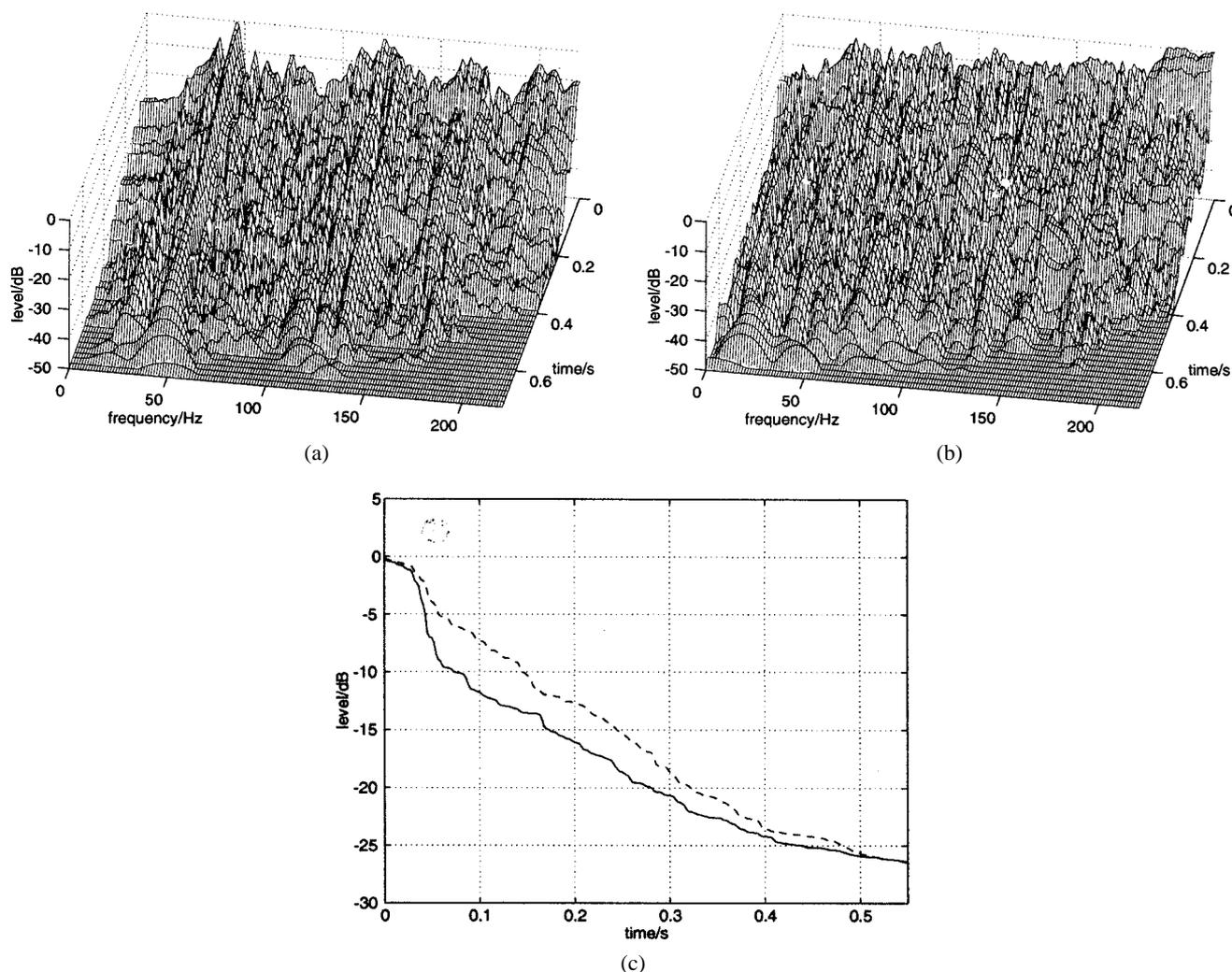


Fig. 13. Case K, modal equalization of room response analyzed in Fig. 8. (a) Original response. (b) After mode equalization. (c) Backward integrated energy decay for original (---) and equalized (—) responses.

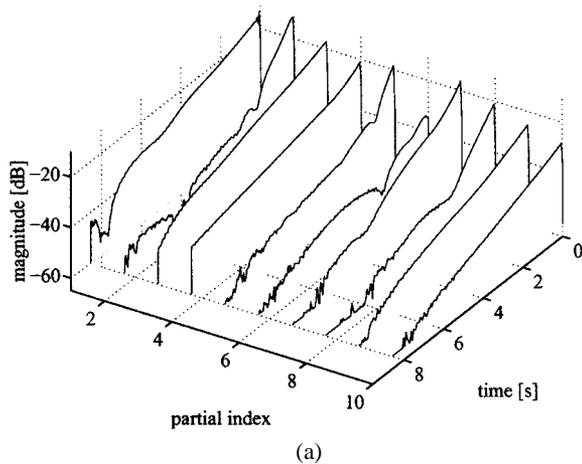
and the estimated amplitude envelope curves for the first 10 harmonics  $f_i$  of a D4 ( $f_0 \approx 298$  Hz) piano tone. The analysis was performed using FZ-ARMA modeling with the parameters  $f_m = f_i$ ,  $K_{\text{zoom}} = 600$ , ARMA (6, 6), and ARMA (12, 12). Order (6, 6) is able to capture the main structure of temporal evolution in partial decays, but the higher order (12, 12) helps with higher precision modeling because for a three-string piano note the degrees of freedom correspond theoretically to  $3 \times 2 \times 2 = 12$  poles.

The string models thus obtained can be used directly for source-filter (= subtractive) synthesis, whereby the total filter order ( $N = P$ ) for single string synthesis varies from 50 to 500, depending on the fundamental frequency and the desired quality of synthesis. Since digital waveguide models [54] are attractive from a computational efficiency point of view, it would be useful to map the modal parameters of the partials to a loop filter of a digital waveguide. To gain advantage from the waveguide formulation, a relatively low-order loop filter is needed which approximates the common properties of the separate modes, thus leading to a nonlinear optimization task.

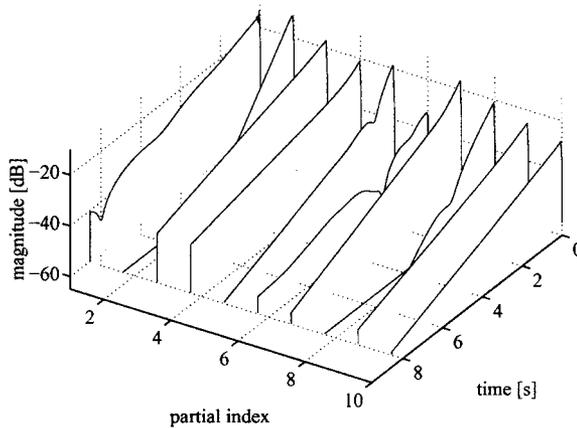
#### 4.4 Modeling of Bell Sounds

As another musical instrument, case N, the analysis and modeling of bell sounds are presented. A characteristic

feature of bell sounds is that they are composed of an inharmonic set of partials [2], such as the one described by the magnitude spectrum shown in Fig. 16(b). Each partial is a decaying sinusoid that, in a closer inspection [Fig. 16(c)], turns out to be a pair or a group of modes located very closely in frequency. This leads to perceptually noticeable beating. In this case the modal group consists primarily of two modes with a frequency difference of about 2.5 Hz.

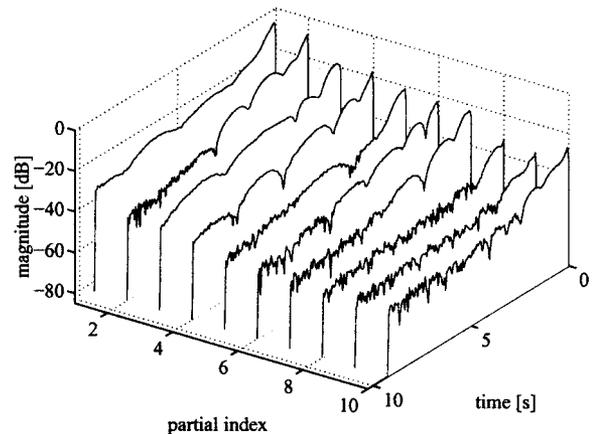


(a)

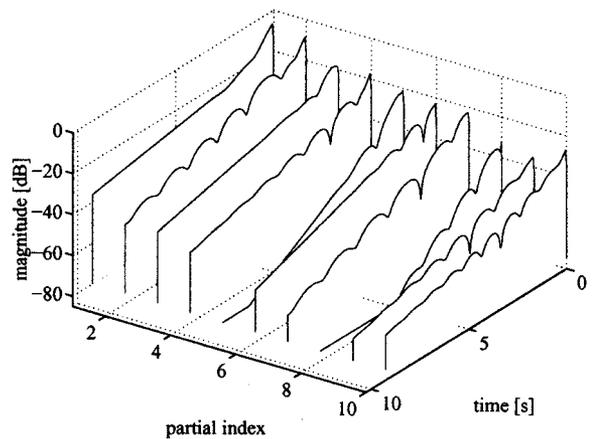


(b)

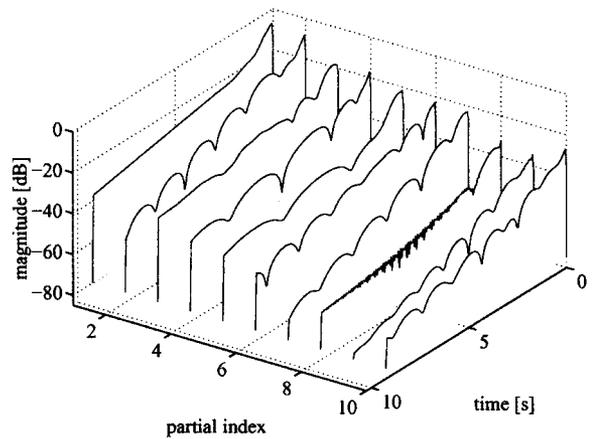
Fig. 14. Case L, analysis of guitar tone partials. (a) Envelopes (in subbands) of first 10 partials of A2 guitar tone. (b) Corresponding partial envelope estimates using FZ-ARMA (6, 6) modeling with  $K_{\text{zoom}} = 600$ .



(a)



(b)



(c)

Fig. 15. Case M, analysis of piano tone partials. (a) Envelopes (in subbands) of first 10 partials of D4 piano tone. (b) Corresponding partial envelope estimates using FZ-ARMA (6, 6) modeling. (c) FZ-ARMA (12, 12) modeling.  $K_{\text{zoom}} = 600$ .

FZ-ARMA is an excellent method for analyzing the modal groups of bell sounds. Fig. 17 shows the envelope match obtained with three different FZ-ARMA orders for the 1310-Hz modal group. The zooming factor  $K_{zoom}$  is 400 in each case. In Fig. 17(a) the orders are  $N = 0$  and  $P = 4$ . Two pole pairs should in principle be sufficient for a double mode, but this all-pole (AR) case with  $N = 0$  does not allow proper phase matching, and thus the over-

all match remains poor.

For ARMA orders  $N = 40$  and  $P = 4$  in Fig. 17(b) the relatively high number of zeros used allows for a good match with just two pole pairs. The same can be achieved in Fig. 17(c) with orders  $N = 2$  to  $P = 6$ , that is, by adding an extra pole pair and keeping the number of zeros minimal. For all resonances up to 10 kHz for this bell sound, filter orders of  $N = 2-4$  and  $P = 4-6$  are sufficient for

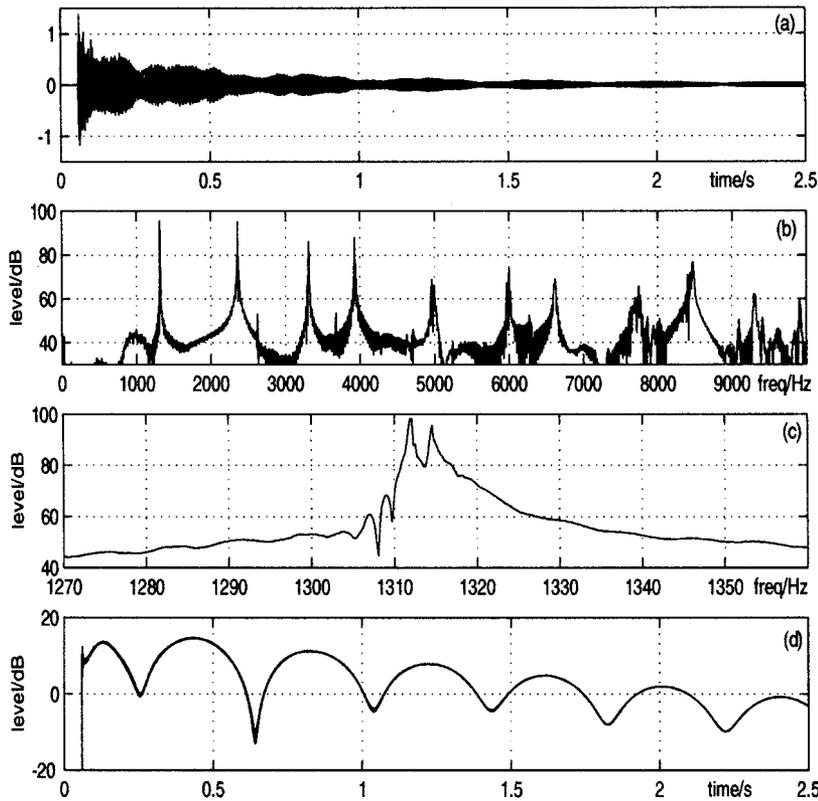


Fig. 16. Acoustical behavior of small bell. (a) Recorded time-domain signal. (b) Magnitude spectrum. (c) Magnitude spectrum in modal region around 1310 Hz. (d) Decay envelope of 1310-Hz modal group.

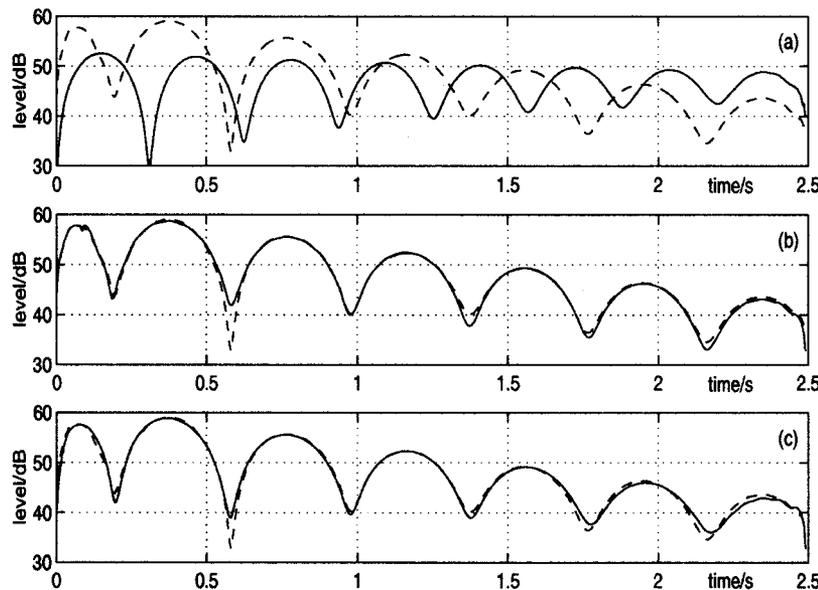


Fig. 17. Case N, analysis and modeling of beating envelope in bell sound mode with different FZ-ARMA model orders. (a) ARMA (4, 0). (b) ARMA (4, 40). (c) ARMA (6, 2).

good modal decay matching so that a parallel filter, composed of modal group filters with a total order of about  $N = 40$  and  $P = 50$ , can implement an efficient and high-quality synthesizer for the bell sound for a sampling rate of 22 050 Hz.

## 5 DISCUSSION AND CONCLUSIONS

In this paper we studied the modeling of acoustic and audio system responses that exhibit resonant and reverberant properties. ARMA modeling techniques were investigated in particular to obtain efficient pole-zero filters. Such modeling, if accurate enough and computationally inexpensive, finds applications in solving many audio-oriented problems.

The first part of the paper is a nontheoretical overview of the AR and ARMA modeling methods to demonstrate their inherent properties and limitations.

A specific interest of this study has been the modeling methods that can yield a good temporal match to a given target response and high frequency resolution, often at the same time. Based on earlier studies, primarily on applying Prony's method to subbands, we showed that frequency-zooming ARMA (FZ-ARMA) based on the Steiglitz-McBride iteration is a powerful technique for high-resolution modeling in subbands. Simulation examples demonstrate the ability of this approach to model complex modal and reverberant behaviors.

These methods were then used to solve problems of practical interest in audio applications, covering room impulse response modeling, inverse modeling for equalization of loudspeaker-room responses, and modeling of musical instrument sounds.

Many general and problem-specific questions remain. Robust selection of model orders, stability and numerical sensitivity, efficient and robust implementation of the filters obtained for subbands or full bands, adaptive formulation and on-line calibration of the models, as well as finding new applications call for more investigation.<sup>12</sup>

## 6 ACKNOWLEDGMENT

This study is part of the VÄRE technology program, project TAKU (Control of Closed Space Acoustics), funded by Tekes (National Technology Agency). The work of Paulo Esquef has been financed by the Brazilian National Council for Scientific and Technological Development (CNPq) and the Academy of Finland, projects "Sound Source Modeling" and "Technology for Audio and Speech Processing."

## 7 REFERENCES

- [1] A. D. Pierce, *Acoustics, An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY, 1989).
- [2] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, New York, 1991).

[3] M. Colloms, *High Performance Loudspeakers*, 5th ed. (Wiley, New York, 1997).

[4] S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. I, *Estimation Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

[5] A. V. Oppenheim, A. Willsky, and I. Young, *Signals and Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[6] S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Upper Saddle River, NJ, 1996).

[7] M. H. Hayes, *Statistical Digital Signal Processing and Modeling* (Wiley, New York, 1996).

[8] MathWorks, Inc., *MATLAB Signal Processing Toolbox, User's Guide* (2001).

[9] MathWorks, Inc., *MATLAB System Identification Toolbox, User's Guide* (2001).

[10] L. B. Jackson, *Digital Filters and Signal Processing* (Kluwer Academic, Boston, MA, 1989).

[11] K. Steiglitz, *A Digital Signal Processing Primer* (Addison-Wesley, Menlo Park, CA, 1996).

[12] Y. Haneda, S. Makino, and Y. Kaneda, "Common Acoustical Pole and Zero Modeling of Room Transfer Functions," *IEEE Trans. Speech Audio Process. (SAP)*, vol. 2, pp. 320–328 (1992 Apr.).

[13] M. Karjalainen, P. Antsallo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of Modal Decay Parameters from Noisy Response Measurements," *J. Audio Eng. Soc.*, vol. 50, pp. 867–878 (2002 Nov.).

[14] J. D. Bunton and R. H. Small, "Cumulative Spectra, Tone Bursts, and Apodization," *J. Audio Eng. Soc.*, vol. 30, pp. 386–395 (1982 June).

[15] J. Makhoul, "Spectral Linear Prediction: Properties and Applications," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 23, pp. 283–296 (1975).

[16] 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. on Acoustics, Speech, and Signal Processing*, vol. 3 (1989), pp. 2053–2056.

[17] M. Schonle, N. Fliege, and U. Zoelzer, "Parametric Approximation of Room Impulse Responses by Multirate Systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1 (Minneapolis, MN, 1993 Apr.), pp. 153–156.

[18] J. Valakas, J. Mourjopoulos, and M. Paraskevas, "Pole-Zero Loudspeaker Response Analysis by Band-splitting and Prony Modeling," presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44, p. 642 (1996 July/Aug.), preprint 4214.

[19] J. D. Markel and J. H. G. Gray, *Linear Prediction of Speech Signals* (Springer, Berlin, 1976).

[20] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ, 1978).

[21] T. W. Parks and C. S. Burrus, *Digital Filter Design* (Wiley, New York, 1987).

[22] L. Ljung, *System Identification—Theory for the User* (Prentice-Hall, Upper Saddle River, NJ, 1999).

[23] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi, "Frequency-Warped

<sup>12</sup>MATLAB code, examples, and sound demonstrations are available at [www.acoustics.hut.fi/demos/AR\\_ARMA/](http://www.acoustics.hut.fi/demos/AR_ARMA/)

Signal Processing for Audio Applications,” *J. Audio Eng. Soc.*, vol. 48, pp. 1011–1031 (2000 Nov.).

[24] T. Paatero and M. Karjalainen, “Kautz Filters and Generalized Frequency Resolution—Theory and Audio Applications,” presented at the 110th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 49, p. 541 (2001 June), preprint 5378.

[25] L. W. P. Biscainho, P. S. Diniz, and P. A. A. Esquef, “ARMA Processes in Sub-Bands with Application to Audio Restoration,” in *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS 2001)*, vol. 2 (Sydney, Australia, 2001 May), pp. 157–160.

[26] J. R. Hopgood and P. J. W. Rayner, “A Probabilistic Framework for Subband Autoregressive Models Applied to Room Acoustics,” in *Proc. IEEE 11th Signal Proc. Workshop on Statistical Signal Processing* (2001), pp. 492–495.

[27] J. A. Moorer, “Signal Processing Aspects of Computer Music: A Survey,” *Proc. IEEE*, vol. 65, pp. 1108–1137 (1977 Aug.).

[28] V. Välimäki and T. Tolonen, “Development and Calibration of a Guitar Synthesizer,” *J. Audio Eng. Soc.*, vol. 46, pp. 766–778 (1998 Sept.).

[29] M. W. Macon, A. McCree, W. M. Lai, and V. Viswanathan, “Efficient Analysis/Synthesis of Percussion Musical Instrument Sound Using an All-Pole Model,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 6 (1998), pp. 3589–3592.

[30] P. P. Vaidyanathan, *Multirate Systems and Filter Banks* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

[31] T. Tolonen, V. Välimäki, and M. Karjalainen, “Modeling of Tension Modulation Nonlinearity in Plucked Strings,” *IEEE Trans. Speech, Audio Process.*, vol. 8, pp. 300–310 (2000 May).

[32] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, “AR/ARMA Analysis and Modeling of Modes in Resonant and Reverberant Systems,” presented at the 112th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 50, pp. 519, 520 (2002 June), preprint 5590.

[33] M. Schonle, N. Fliege, and U. Zoelzer, “Parametric Approximation of Room Impulse Responses Based on Wavelet Decomposition,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA ’93)*, (New Paltz, NY, 1993 Oct.), pp. 68–71.

[34] M. R. Schroeder, “Statistical Parameters of the Frequency Response Curves of Large Rooms,” *J. Audio Eng. Soc.*, vol. 35, pp. 299–306 (1987 May).

[35] T. Paatero and M. Karjalainen, “New Digital Filter Techniques for Room Response Modeling,” in *Proc. AES 21st Conf.* (St. Petersburg, Russia, 2002 June).

[36] J. Mourjopoulos, “Digital Equalization Methods for Audio Systems,” presented at the 84th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 36, p. 384 (1988 May), preprint 2598.

[37] J. Kuriyama and Y. Furukawa, “Adaptive Loudspeaker System,” *J. Audio Eng. Soc. (Engineering Reports)*, vol. 37, pp. 919–926 (1989 Nov.).

[38] S. J. Elliott and P. A. Nelson, “Multiple-Point Equalization in a Room Using Adaptive Digital Filters,” *J.*

*Audio Eng. Soc.*, vol. 37, pp. 899–907 (1989 Nov.).

[39] J. Mourjopoulos and M. A. Paraskevas, “Pole and Zero Modelling of Room Transfer Functions,” *J. Sound Vibr.*, vol. 146, pp. 281–302 (1994 Nov.).

[40] R. G. Greenfield, “Application of Digital Techniques to Loudspeaker Equalization,” PhD thesis, Department of Electronic Systems Engineering, University of Essex, UK (1991 Sept.).

[41] R. Genereux, “Adaptive Filters for Loudspeakers and Rooms,” presented at the 93rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, pp. 1044, 1045 (1992 Dec.), preprint 3375.

[42] F. X. Y. Gao and W. M. Snelgrove, “Adaptive Linearization of a Loudspeaker,” presented at the 93rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, p. 1045 (1992 Dec.), preprint 3377.

[43] J. Mourjopoulos, “Digital Equalization of Room Acoustics,” *J. Audio Eng. Soc.*, vol. 42, pp. 884–900 (1994 Nov.).

[44] G. Karachalios, D. Tsoukalas, and J. Mourjopoulos, “Multiband Analysis and Equalization of Loudspeaker Responses,” presented at the 98th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 43, p. 396 (1995 May), preprint 3975.

[45] S. Salamouris, K. Politopoulos, V. Tsakiris, and J. Mourjopoulos, “Digital System for Loudspeaker and Room Equalization,” presented at the 98th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 43, p. 396 (1995 May), preprint 3976.

[46] Y. Haneda, S. Makino, and Y. Kaneda, “Multiple-Point Equalization of Room Transfer Functions by Using Common Acoustical Poles,” *IEEE Trans. Speech Audio Process. (SAP)*, vol. 5, pp. 325–333 (1997 July).

[47] M. Karjalainen, E. Piirilä, A. Järvinen, and J. Huopaniemi, “Comparison of Loudspeaker Response Equalization Methods Based on DSP Techniques,” *J. Audio Eng. Soc.*, vol. 47, pp. 14–31 (1999 Jan./Feb.).

[48] F. Fontana, “Common Pole Equalization of Small Rooms Using a Two-Step Real-Time Digital Equalizer,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA ’99)* (1999), pp. 195–198.

[49] M. Tyril, J. A. Pedersen, and P. Rubak, “Digital Filters for Low-Frequency Equalization,” *J. Audio Eng. Soc. (Engineering Reports)*, vol. 49, pp. 36–43 (2001 Jan./Feb.).

[50] A. V. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Välimäki, “Low-Frequency Modal Equalization of Loudspeaker–Room Responses,” presented at the 111th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 49, p. 1232 (2001 Dec.), preprint 5480.

[51] A. V. Mäkivirta and C. Anet, “A Survey Study In-Situ Stereo and Multi-Channel Monitoring Conditions,” presented at the 111th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 49, p. 1236 (2001 Dec.), preprint 5496.

[52] AES Technical Committee on Multichannel and Binaural Audio Technology (TC-MBAT), “Multichannel

Surround Sound Systems and Operations,” Tech. Doc., version 1.5, Audio Engineering Society, New York (2001).

[53] EBU Doc. Tech. 3276-1998, “Listening Condition for the Assessment of Sound Programme Material: Monophonic and Two-Channel Stereophonic,” 2nd ed., EBU (1998).

[54] J. O. Smith, “Physical Modeling Using Digital Waveguides,” *Computer Music J.*, vol. 16, no. 4, pp. 74–91 (1992).

[55] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy, “Physical Modeling of Plucked String Instruments with Application to Real-Time Sound Synthesis,” *J. Audio Eng. Soc.*, vol. 44, pp. 331–353 (1996 May).

[56] J. O. Smith, “Techniques for Digital Filter Design and System Identification with Application to the Violin,” PhD thesis Rep. STAN-M-14, CCRMA, Department of Music, Stanford University, Stanford, CA (1983 June).

[57] M. Sandler, “Algorithm for High Precision Root Finding from High Order LPC Models,” *IEE Proc. I on Communications, Speech and Vision*, vol. 138, pp. 596–602 (1991 Dec.).

[58] M. Sandler, “New Synthesis Techniques for High-Order All-Pole Filters,” *IEE Proc. G, Circuits, Devices and Systems*, vol. 139, pp. 9–16 (1992 Feb.).

[59] J. Laroche and J. L. Meillier, “Multichannel Excitation/Filter Modeling of Percussive Sounds with Application to the Piano,” *IEEE Trans. Speech Audio Process.*, vol. 2, pp. 329–344 (1994 Apr.).

[60] J. Mackenzie, I. Kale, and G. D. Cain, “Applying Balanced Model Truncation to Sound Analysis/Synthesis Models,” in *Proc. Int. Computer Music Conf.*, (1995 Sept.), pp. 400–403.

[61] M. Karjalainen, V. Välimäki, and T. Tolonen, “Plucked-String Models, from the Karplus–Strong Algorithm to Digital Waveguides and Beyond,” *Computer Music J.*, vol. 22, no. 3, pp. 17–32 (1998).

---

The biographies of Matti Karjalainen, Paulo A. A. Esquef, and Vesa Välimäki were published in the 2002 April issue of the *Journal*.

The biography of Poju Antsalu was published in the 2002 November issue of the *Journal*.

The biography of Aki Mäkipirta was published in the 2002 July/August issue of the *Journal*.