System Measurement and Identification Using Pseudorandom Filtered Noise and Music Sequences*

M. O. J. HAWKSFORD, *AES Fellow*

Centre for Audio Research and Engineering, University of Essex, Colchester, C043SQ, UK

System measurement using pseudorandom filtered noise and music sequences is investigated. A single-pass technique is used to evaluate simultaneously the transfer function and the spectral-domain signal-to-distortion ratio that is applicable to amplifiers, signal processors, digital-to-analog converters, loudspeakers, and perceptual coders. The technique is extended to include a simplified Volterra model expressed as a power series and linear filter bank where for compliant systems, nonlinear distortion can be estimated for an arbitrary excitation without a need for remeasurement.

0 INTRODUCTION

This paper explores methods of determining a system's transfer function using pseudorandom noise applied in a single-pass process and builds on earlier work by Borish and Angell [1] and, later, Vanderkooy and Rife [2]. Both linear and nonlinear distortion is considered, and a simplified method of system identification is introduced that models a class of system based on a Taylor series, but where each power term in the series is filtered by a unique transfer function. The nonlinear kernels of this model form a subset of a full Volterra model and are extracted here using concatenated finite noise sequences, a method that may be considered the dual of the time-domain spectrometry (TDS) approach of Farina [3]. Although this model is not universal, it can be applied to a range of audio systems and forms a bridge between input-specific measurements and formal methods of identification [4]–[6].

As a further extension, techniques are presented using comb-filtered pseudorandom noise, which allows the simultaneous estimation of both linear and nonlinear distortion by determining the distortion residue falling within the spectral nulls of the excitation. In addition to noise, alternative audio signals can be substituted to enable the relationship between signal and distortion to be explored. Thus by extracting the distortion waveform and using classic block transform analysis the measurement technique can be extended to include systems exhibiting dynamic and perceptually motivated nonlinearity. Finally a graph depicting both the magnitude transfer function and an estimate of the linear error function is presented that offers a more holistic picture of linear system performance. This graph preserves the small differences in frequency response that are often lost because of limited display resolution and system noise. It also expresses the linear dynamic range (LDR) based upon the target signal and residue resulting from linear system error. Although there are caveats to this approach, additional insight into system behavior may be gained where, as an example, a high-performance CD player is assessed.

The exploitation of noise signals has had a long history in the field of system measurement [7], [8]. For example, continuous Gaussian white noise has been used to evaluate a system's transfer function and has proved effective because of its persistent nature and improved measurement signal-to-noise ratio (SNR). More recently maximum-length binary sequences (MLSs) [1], [2] have been used especially for loudspeaker and acoustic measurements. By selecting a periodic MLS of appropriate length to minimize time aliasing distortion [2], a system's periodic impulse response can be calculated directly by circular cross correlation of the measured sequence and the excitation sequence [1], [2]. An interesting observation when performing MLS-based measurements on a nonlinear system is that the resultant impulse response includes a broad distribution of minor impulses [9]. Hence part of the motivation for the present work is to exploit this phenomenon and to construct a simplified nonlinear Volterra model of the system being measured, derived from data extracted using noise sequences.

*Presented at the 114th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 2003 March 22–25; revised 2005 January 14 and February 9.

The techniques described can achieve state-of-the art measurements that compare favorably with MLS methods and multitone methods. Indeed by suitable spectral weighting of the excitation, some of the methods presented here may be considered a generalization of multitone testing. The principal limitations, as with most systems, lie with the quality of the converters used; however, the core processing easily exceeds the resolution of practical audio equipment. The system was developed from a need to explore the performance limits of high-performance CD players, where the aim was to determine the limits of both linear and nonlinear performance, preferably using a single-pass measurement. In this application it was also desirable to have a noise excitation with a uniform probability density function in order to fully exercise the converters; thus binary MLS was rejected (although it is recognized that there are filtering techniques that can address this limitation). Also, it was recognized that Fourier transform techniques were just as effective as the Hadamard transform often used with MLS. In addition, the methods have also been used to assess algorithms in desktop audio editors, such as sample rate converters, of which an example is given in Section 4.3 of course in these applications there is no limitation imposed by converters. The nature of the techniques applied here is that they are easily adapted where, for example, the decision to include fine frequency filtering together with broad-band excitation was seen as pivotal. Such an approach can be configured to push a system under test to its performance boundary so that extremely complex patterns of distortion can be exposed and where, to introduce greater reality, even music-derived signals may be employed. The work also allowed several research threads to be merged, including a long-term personal interest in errors related to linear distortion and to nonlinear modeling. For example, the measurement system can be adapted both to excitation-specific system measurements and to a simplified method of Volterra identification, where it acts as a bridge between the two approaches. It also became evident that once spectral-domain filtering was incorporated to facilitate the segregation of excitation and the resulting distortion, then the problem of measuring both time-varying and non-time-varying systems should be addressed. As a result, the paper offers a contribution to both approaches, where by translating the comb-filtering methodology into the z domain it becomes possible to construct a system using short-term spectral analysis to track dynamic distortion as encountered, for example, with perceptual codecs.

In the measurement procedures described in this study, a rectangular windowed noise (or in some cases music) segment constitutes a frame,[1] with several frames then concatenated to approximate a continuous sequence. An individual frame may also be equalized to have constant-magnitude spectrum but random phase over the length $N = 2^K$ ($2^K - 1$ for MLS). Here K is a positive integer for compatibility with fast Fourier transform (FFT) procedures. Thus most measurement advantages of MLS are retained. Parallels can also be drawn with multi-sine-wave testing [10], [11] since a repetitive noise sequence constitutes a multitone signal, where the fundamental frequency is the frame repetition rate with harmonics forming the tones. Consequently if a noise frame is equalized for a flat-magnitude spectrum, then the multiple tones are of equal amplitude but with random phase relationships.

All processing described in this paper was written in Matlab[2] running on a PC interfaced to high-quality converters. The paper therefore adopts Matlab notation to describe vector operations. Also, to evaluate DVD/CD players (see example measurement, Section 5), the test signals can be burnt to CD/DVD, thus eliminating the need for dedicated test equipment, with further benefits accrued in terms of convenience as all tests employ a single-pass procedure.

The study commences by describing the noise sequence, its equalization and transfer function derivation. Consideration is also given to the formation of a composite test sequence and factors pertaining to the selection of frame length. In all measurement variants discussed it is assumed that analog-to-digital converter (ADC) and digital-to-analog converter (DAC) sampling rates are synchronized as this is critical to proper transform analysis in relation to frame size. Nonsynchronous operation is not considered in this study since both incurred processing errors and remedial windowing artifacts reduce measurement precision. Where a CD/DVD player is used either as an excitation source or for its evaluation, the ADC is slaved to the player sampling rate via the standard Sony/Philips digital interface (S/PDIF).

1 LINEAR SYSTEM IDENTIFICATION USING PSEUDORANDOM NOISE

The core technique exploited this study to measure a system's transfer function is based on a repetitive equalized noise sequence (that is, pseudorandom noise), where a noise sequence must be generated with a duration greater than the time over which a system's impulse response $h(t)$ remains significant. The noise sequence is defined in discrete time, where Nyquist sampling theory determines the sampling rate as a function of bandwidth. Consequently measurement accuracy is bounded by both time [2] and frequency [12] aliasing distortion. The noise sequence is concatenated to form repetitive frames with no interframe guard bands. It then follows that to extract spectral information, only sampling-rate synchronization is required; exact frame synchronization, although beneficial, is not mandatory, provided the frame size is known, as the transforms used are circular. Consequently a sample-rate synchronized ADC captures the output response of the system being tested and the frame detection achieved both by counting the frames of 2^K samples and using a synchronization preamble embedded in the test sequence.

Consider a noise vector noise(n) with rectangular probability density function, generated over $N = 2^K$ samples,

[1]A frame is defined here as a finite set of uniform samples represented as a vector.

[2]Matlab is a trade name of Mathworks, Inc.

where K is a positive integer and n is the vector [1:N]. Expressed in Matlab notation,

$$\text{noise}(n) = \text{rand}(1, N). \tag{1}$$

The frequency-domain noise sequence noisef(n) is calculated using a one-dimensional FFT. Thus if fft(noise(n)) is the length-N discrete Fourier transform of the sequence noise(n), where N is the length of the sequence noise(n), then

$$\text{noisef}(n) = \text{fft}(\text{noise}(n)). \tag{2}$$

A time-domain excitation sequence test(n) with constant-magnitude spectrum but random phase is then determined using spectral normalization and the inverse Fourier transform, where[3]

$$\text{test}(n) = \text{real}(\text{ifft}(\text{noisef}(n)./(\gamma + \text{abs}(\text{noisef}(n))))). \tag{3}$$

Although processing using the inverse Fourier transform should in this application return only real numbers, computational errors result in small but finite imaginary terms, which is not the norm for time-domain sampled data and is unacceptable when writing a wav file. Hence only the real part of the transform is selected both here and in later inverse transform operations. In addition, the real function also halves the vector storage requirement as the small nonzero imaginary elements are deleted. Also, a constant γ (say 10^{-12}) is introduced to eliminate small-number division anomalies in the spectral normalization process. Alternatively this potential problem can be avoided completely by using a complex exponential function testf(n) with random phase to guarantee an exactly constant magnitude spectrum,

$$\text{testf}(n) = \exp(\text{i}^*\text{angle}(\text{fft}(\text{noise}(n)))) \tag{4}$$

that is, $|\text{testf}(n)| = 1$. The corresponding time-domain vector test(n) then follows as

$$\text{test}(n) = \text{real}(\text{ifft}(\text{testf}(n))). \tag{5}$$

In practice, because the test sequences have to be generated only once, it is prudent to sift a number of computed examples in order to seek a sequence with low crest factor such that the measurement SNR can be enhanced. A composite repetitive excitation pattern of test(n) is then constructed, as shown in Fig. 1, which includes both a zero pulse preamble and an embedded synchronization sequence defined as [0 0 0 . . . 0 0 0 0 1 1 −1 −1 0 0 0 0]. Hence by using cross-correlation-based detection the commencement of the recovered test sequence can be detected to sample accuracy. Finally the composite sequence is peak amplitude normalized, quantized to the required bit depth, and converted to a two-channel linear pulse code modulation (LPCM) wav file for subsequent outputting to the system under test, typically via a 96-kHz, 24-bit DAC.

The measured data are captured using a sample synchronized ADC, where following frame synchronization a data frame output(n) is extracted. Taking testf(n) from Eq. (4), the complex transfer function TF(n) of the system is then calculated using element-by-element division,

$$\text{TF}(n) = \text{fft}(\text{output}(n))./\text{testf}(n). \tag{6}$$

The magnitude response $M(n)$ and the phase response $P(n)$ then follow as

$$M(n) = \text{abs}(\text{TF}(n)) \quad \text{and} \quad P(n) = \text{angle}(\text{TF}(n)) \tag{7}$$

and the system impulse response $h(n)$ as

$$h(n) = \text{real}(\text{ifft}(\text{TF}(n))). \tag{8}$$

Alternatively, if excess phase information is not required, then the magnitude response of the spectrum may be calculated directly from output(n) since the excitation was normalized to a constant-magnitude spectrum. The minimum-phase impulse response $h_{\min}(n)$ then follows from the Hilbert transform [13],

$$h_{\min}(n) = \text{real}(\text{ifft}(\exp(\text{conj}(\text{hilbert}(\log(\text{abs}(\text{fft}(\text{output}(n))))))))). \tag{9}$$

However, if frame synchronization is not achieved, then because of circularity and repetitive noise frames, the true impulse response can still be derived, but within an arbitrary time shift. A key factor in this process is for the noise frame to exceed the duration of $h(n)$. Otherwise the circular nature of the test procedure allows time aliasing distortion, which is a fundamental and irreversible measurement distortion. With this proviso then, for a linear system, this procedure creates an exact model within the constraints of measurement bandwidth and sampling rate. In the next section the process is extended to include approximate nonlinear identification employing a simplified Volterra model.

2 NONLINEAR SYSTEM IDENTIFICATION USING PSEUDORANDOM NOISE

Farina [3] has reported a TDS-based scheme to identify mildly nonlinear systems in terms of a simplified Volterra model. An alternative measurement procedure is described here using pseudorandom noise similar to that presented in Section 1. The model is appropriate for stationary nonlinear systems with memory where Volterra kernels expressed as impulse responses encapsulate higher order frequency dependence. However, only powers of the input

[3]See Matlab glossary in Appendix A for a definition of operator ".”

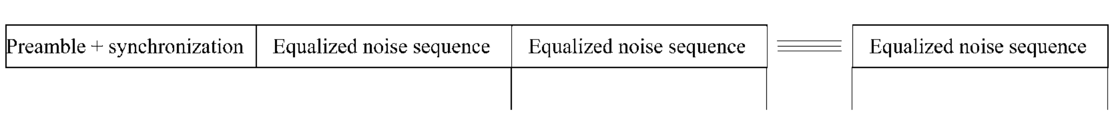


Fig. 1. Test signal structure with preamble and synchronization sequence.

sequence are included whereas cross-product terms inherent in the full Volterra model are ignored. As such the method is positioned between input-specific measurements and a fully populated Volterra model that can predict the output response to a generalized excitation [14]–[16]. Also as a corollary, linearity can be tested as linear and nonlinear responses are segregated.

2.1 Volterra Modeling

The general Volterra model linking input vector $x(n)$ to output vector $y(n)$ is described by M-dimensional convolution,

$$y(l)\Big|_{l=1}^{N} = \sum_{r=1}^{M}\sum_{r=1}^{N-1}\cdots\sum_{r=1}^{N-1} h_r(i_1, i_2, \ldots, i_r) \times x(l-i_1)x(l-i_2)\cdots x(l-i_r) \qquad (10)$$

where N is the memory length of the filters. Consequently the general Volterra model requires a large number of coefficients to populate a multidimensional space as unique impulse responses are associated with all the convolutional combinations of power and cross-product terms. However, in the simplified representation the only convolutions included are those associated with powers of the input sequence.

Consider a nonlinearity where output vector $y(n)$ is related to input vector $x(n)$ by a power series of order M,

$$\begin{aligned} y(n) &= a_0 + a_1x(n) + a_2x(n).\hat{}2 + \cdots + a_rx(n).\hat{}r + \cdots + a_Mx(n).\hat{}M \\ &= [a_0\ a_1\ a_2 \ldots a_M].*[1\ x(1)\ x(2)^2 \ldots x(n)^r \ldots x(n)^M] \\ &= [a].*[1\ x(1)\ x(2)^2 \ldots x(n)^r \ldots x(n)^M]. \end{aligned} \qquad (11)$$

To identify this memoryless system fully, only the M coefficients $[a]$ need to be determined. However, for the simplified Volterra model with memory the $[a]$ coefficients associated with each term in the power series translate to a set of M impulse responses $[h(n)]$, reducing the M-dimensional convolution[4] in Eq. (10) to just

$$\begin{aligned} y(n) = h_0 &+ h_1(n) \otimes x(n) + h_2(n) \otimes (x(n).\hat{}2) + \cdots \\ &+ h_r(n) \otimes (x(n).\hat{}r) + \cdots \\ &+ h_M(n) \otimes (x(n).\hat{}M). \end{aligned} \qquad (12)$$

In Eq. (12) h_0 is the dc term and $h_1(n)$ is the linear system impulse response, while for $r = 2, \ldots, M$, $h_r(n)$ describes the respective impulse responses relating to the power terms $x(n).\hat{}r$. Because the Volterra model described by Eq. (12) contains M impulse responses, M independent noise sequences are required in the identification procedure, although vectors are transformed into the frequency domain to allow simpler element-by-element multiplication rather than time-domain convolution.

2.1.1 *Vector and Transform Notation*

For an M-dimensional system $x_r^s(n)$ represents an input vector r where each element is raised to the power s, $y_r(n)$ is output vector r, and $h_r(n)$ is an impulse response r. The corresponding Fourier transforms Y_r, $X_{r,s}$, H_r are then defined,

$$Y_r = \mathrm{fft}(y_r(n)), \qquad X_{r,s} = \mathrm{fft}(x_r^s(n)), \qquad H_r = \mathrm{fft}(h_r(n)).$$

In the following analysis vectors are transformed between time and frequency domains to transmute convolution to element-by-element multiplication. Consider M Fourier-transformed vectors Y_r derived from M uncorrelated excitation noise vectors $X_{r,1}$ and applied successively to Eq. (12) to form M equations, that is,

$$\begin{aligned} Y_1 &= H_0 + H_1.*X_{1,1} + H_2.*X_{1,2} + \cdots + H_r.*X_{1,r} + \cdots + H_M.*X_{1,M} \\ Y_2 &= H_0 + H_1.*X_{2,1} + H_2.*X_{2,2} + \cdots + H_r.*X_{2,r} + \cdots + H_M.*X_{2,M} \\ &\vdots \\ Y_M &= H_0 + H_1.*X_{M,1} + H_2.*X_{M,2} + \cdots + H_r.*X_{M,r} + \cdots + H_M.*X_{M,M}. \end{aligned}$$

Rewriting in matrix form,

$$\begin{bmatrix} Y_1 - H_0 \\ Y_2 - H_0 \\ \vdots \\ Y_M - H_0 \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,M} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,M} \\ \vdots & \vdots & & \vdots \\ X_{M,1} & X_{M,2} & \cdots & X_{M,M} \end{bmatrix} .* \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_M \end{bmatrix} \qquad (13)$$

that is,

$$[Y] = [X].*[H]. \qquad (14)$$

In Eq. (13) H_0 is the output dc offset. It is measured when the input signal is zero and the system quiescent. To determine impulse responses $[H]$, define $[Z] = [X]^{-1}$ (see Appendix B.1 for inversion), whereby the decoding matrix equation expressed in terms of M measured output vectors becomes

$$[H] = [X]^{-1}.*[Y] = [Z].*[Y]. \qquad (15)$$

Eq. (15) describes an input-specific decoding key, where $[Z]$ is related uniquely to the set of M noise vectors and has to be calculated only once. This simplifies computation, since a typical $[M, M, N]$ matrix for $M = 8$ and $N = 2^{14}$ contains 2^{20} complex elements. To complete the analysis the set of Volterra impulse responses $[h]$ follow from the inverse fast Fourier transform of matrix $[H]$, where

$$[h] = \mathrm{real}(\mathrm{ifft}([H])). \qquad (16)$$

2.2 Test Sequence Generation

In performing a system measurement it is critical for each of the M noise vectors to have the same relative level. To facilitate this requirement, a composite signal is constructed where each noise vector is repeated four times to

[4]Although not a Matlab operator, the symbol $\otimes$ represents circular convolution.

form a subframe, then all M subframes are concatenated into a single sequence. Subframes containing repeated sequences allow convergence to a pseudoperiodic output signal and introduce a margin against subframe misalignment within the decoder. Repeated frames also enable noise averaging to be applied to improve SNR. A preamble and synchronization sequence is then added similar to that described in Section 1 to facilitate demultiplexing of the M sequential system responses. Consequently all M sequences are processed almost simultaneously in a single measurement so differential gain errors are eliminated.

2.3 Volterra Modeling Validation

To validate the Volterra modeling scheme the computational process is divided into three routines (discussed in Appendix B.2 and B.3) and then applied to two nonlinear examples.

- An M-vector composite test sequence is generated and the three-dimensional decoding matrix $[X]$ inverted to $[Z]$ as a one-off calculation.
- Two example simulations are performed on a stationary nonlinearity, with and without memory.
- Output data are analyzed and Volterra responses computed using Eq. (15).

1) *Nonlinearity without Memory.* The first example system employs just a power series as described by Eq. (11), where the excitation is processed sequentially, sample by sample. The coefficient matrix $[a]$ of the series is selected arbitrarily and the performance of the decoding algorithm evaluated by comparing the amplitude of the Volterra frequency-domain responses to $[a]$. For this memoryless case each Volterra response has constant amplitude, assuming no measurement channel filtering. The selected coefficient values for $M = 8$ are

$$[a] \equiv [0 \quad 1 \quad 0.001 \quad 0.05 \quad 0.0002 \quad 0.02 \quad 0.0005 \quad 0.05 \quad 0.001].$$

2) *Nonlinearity with Memory.* The second example, depicted in Fig. 2 for $M = 8$, adds a set of linear low-pass output filters applied to each power term of the nonlinear series described by Eq. (12). All eight filters have brick-wall responses with respective cutoff frequencies of 16, 14, 12, 10, 8, 6, 4, and 2 kHz.

Volterra analysis was applied to each nonlinear system, where $N = 2^{15}$. For the memoryless case the eight derived Volterra frequency-domain responses are shown in Fig. 3, those for the second example with memory are shown in Fig. 4. Results correspond to theory within the bounds of measurement noise where all responses match those preselected, including correct identification of the eight brick-wall filter responses.

3 NONLINEAR DISTORTION ESTIMATION USING COMB-FILTERED NOISE SEQUENCES

It is known that nonlinear systems when excited by broad-band signals produce complicated spectral patterns of intermodulation distortion [9]. Section 2 exploited this phenomenon in Volterra identification. In this section it is shown that a combination of noise excitation and comb filtering enable distortion and signal to be partially separated, thus allowing simultaneous estimates of both linear and nonlinear distortion. The proposed spectral interleave technique takes inspiration from both Belcher [17] and techniques of multitone testing [10], [11].

3.1 Evaluation of Spectral Interleave Measurement System

The measurement procedure developed in Section 1 is extended to include frequency-domain comb filtering by introducing an interleave-mask function intf0(n). The mask forces regions of zero spectral energy in the excitation sequence and is also used in measurement analysis to segregate signal and distortion. Applying the frequency-domain vector testf(n) with constant magnitude and random phase from Eq. (4), the frequency-normalized and comb-filtered test sequence $\text{test}_{\text{int}}(n)$ becomes

$$\text{test}_{\text{int}}(n) = \text{real}(\text{ifft}(\text{intf0}(n).{*}\text{testf}(n))). \tag{17}$$

Two interleave-mask options were incorporated for frequency domain filtering:

1) Alternating binary sequence . . . 010101 . . . to create a regular pattern of active and zero frequency bins.

2) MLS to create a random pattern of active and zero spectral bins.

To construct a frequency-domain mask that takes proper account of the sampled-data format of the excitation, appropriate spectral symmetry is required about the half-sampling frequency, where if N is the number of samples in the noise sequence,

$$\text{intf0}(1{:}N) = [0 \quad \text{intf0}(1{:}N/2-1) \quad 0 \quad \text{intf0}(N/2-1{:}-1{:}1)]. \tag{18}$$

Hence if mdata(n) is measured data windowed precisely to match the excitation sequence, then applying testf(n) from Eq. (4), the system frequency response GF(n) is extracted as

$$\text{GF}(n) = \text{intf0}(n).{*}[\text{fft}(\text{mdata}(n))./\text{testf}(n)] \tag{19}$$

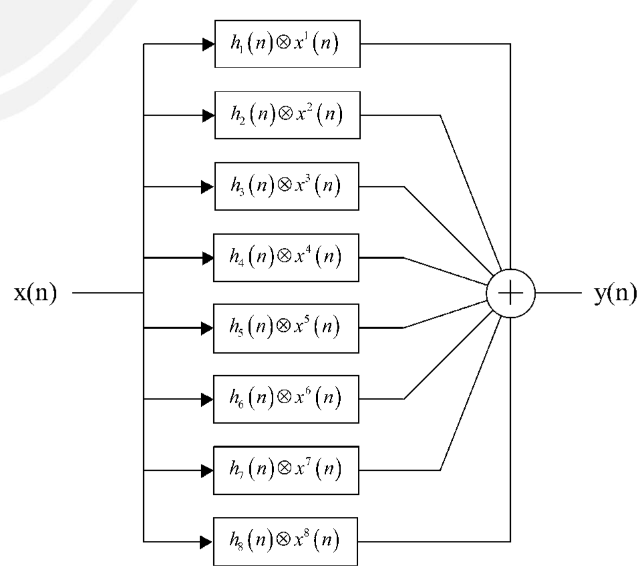


Fig. 2. Simplified Volterra model based on Eq. (12).

while the distortion spectrum DF(n) is obtained using the complementary mask, where

$$\mathrm{DF}(n) = [\mathrm{ones}(1,N) - \mathrm{intf0}(n)].*[\mathrm{fft}(\mathrm{mdata}(n))./\mathrm{testf}(n)]. \tag{20}$$

In practice vector lengths up to $N = 2^{20}$ have been used successfully without experiencing computational problems in the Matlab FFT. This represents an excitation repetition period of approximately 23.77 second at a sampling rate of 44.1 kHz, corresponding to a frequency-bin

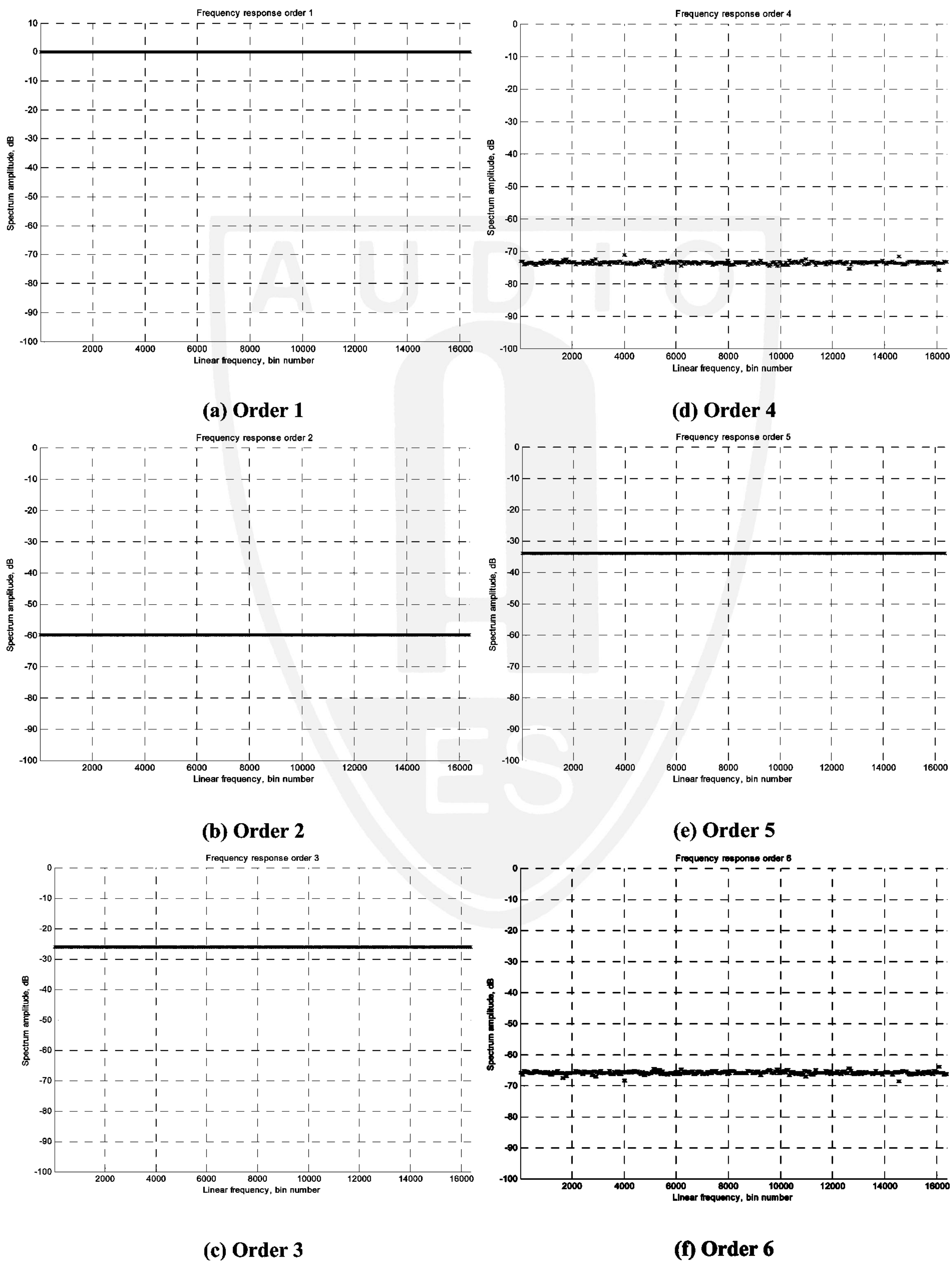


Fig. 3. Volterra frequency-domain responses for nonlinear system without memory.

spacing of 0.042 Hz. Because of the fine frequency resolution achievable, precise sampling rate synchronization is mandatory; otherwise spectral leakage into adjacent bins introduces false estimates of distortion.

3.2 Noise Averaging and Full Spectral Resolution Using Multiple Frames

Because excitation sequences are quasi repetitive, it is straightforward to capture a number of consecutive frames and perform noise averaging to improve the measurement SNR, where the theoretical noise improvement $\mathrm{SNR}(N)|_{\mathrm{imp}}$ for λ averaged frames is

$$\mathrm{SNR}(N)|_{\mathrm{imp}} = 10\log_{10}(\lambda). \tag{21}$$

However, once multiple frames are used, two sets of frames can be constructed using complementary interleave masks, where, for example, four frames are assigned a normal interleave mask while a subsequent four frames are assigned a complementary interleave mask. Hence a single compound measurement sequence enables measured data to be gathered where the respective frequency bins assigned to frequency response and distortion estimation are interchanged. The two sets of data from successive groups of frames can then be segregated temporally and subsequently merged to produce noninterleaved full-resolution spectra for both the transfer function $\mathrm{GF}_{\mathrm{full}}(n)$ and distortion $\mathrm{DF}_{\mathrm{full}}(n)$ as follows.

Let the transfer function and the distortion spectra derived from the first set of measured sequences mdata_1 be

$$\mathrm{GF}_1(n) = \mathrm{intf0}(n).*[\mathrm{fft}(\mathrm{mdata}_1(n))./\mathrm{testf}(n)] \tag{22}$$

$$\mathrm{DF}_1(n) = [\mathrm{ones}(1,N) - \mathrm{intf0}(n)].*[\mathrm{fft}(\mathrm{mdata}_1(n))./\mathrm{testf}(n)] \tag{23}$$

and from the second set of measured sequences mdata_2,

$$\mathrm{GF}_2(n) = [\mathrm{ones}(1,N) - \mathrm{intf0}(n)].*[\mathrm{fft}(\mathrm{mdata}_2(n))./\mathrm{testf}(n)] \tag{24}$$

$$\mathrm{DF}_2(n) = \mathrm{intf0}(n).*[\mathrm{fft}(\mathrm{mdata}_2(n))./\mathrm{testf}(n)]. \tag{25}$$

The full-resolution spectra $\mathrm{GF}_{\mathrm{full}}(n)$ and $\mathrm{DF}_{\mathrm{full}}(n)$ are then calculated,

$$\mathrm{GF}_{\mathrm{full}}(n) = \mathrm{GF}_1(n) + \mathrm{GF}_2(n) \tag{26}$$

$$\mathrm{DF}_{\mathrm{full}}(n) = \mathrm{DF}_1(n) + \mathrm{DF}_2(n). \tag{27}$$

The corresponding non-comb-filtered time-domain periodic sequences $\mathrm{out}_{\mathrm{f}}(n)$ and $\mathrm{dist}_{\mathrm{f}}(n)$ follow from the inverse fast Fourier transform as

$$\mathrm{out}_{\mathrm{f}}(n) = \mathrm{real}(\mathrm{ifft}(\mathrm{GF}_{\mathrm{full}}(n))) \tag{28}$$

$$\mathrm{dist}_{\mathrm{f}}(n) = \mathrm{real}(\mathrm{ifft}(\mathrm{DF}_{\mathrm{full}}(n))). \tag{29}$$

To gain additional insight into this process, including inherent time smearing of both excitation and retrieved distortion sequences which results from comb filtering, Appendix B.4 presents a time-domain description of the measurement process.

3.3 Examples Using Comb-Filter Measurement Procedure

To evaluate the measurement system incorporating comb filtering, three example systems were simulated. In each of these tests the excitation sequence used 44.1-kHz sampling with 16-bit resolution.

3.3.1 Linear Filter

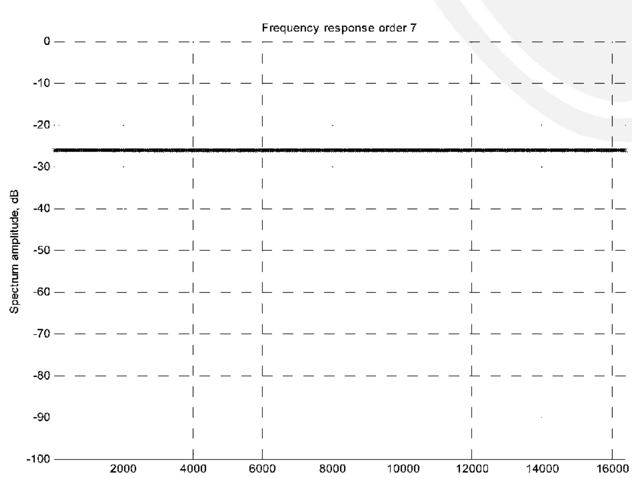
A linear filter was simulated, where the frequency response had two linear segments located above and below 1 kHz, with the attenuation peaking at 4.5 dB. Fig. 5 shows the simulation results, where the correct amplitude response has been obtained and where all "zero bins" contain only quantization noise, thus confirming system linearity.

3.3.2 Nonlinearity, No Filtering

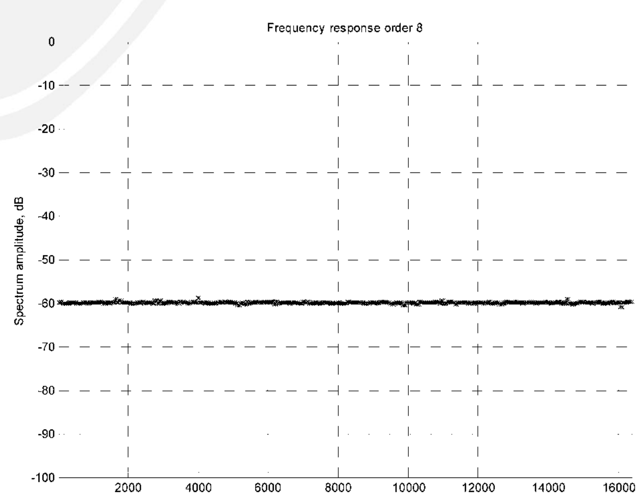
The second example used a memoryless nonlinearity defined by

$$y(n) = x(n) + 0.01x^2(n). \tag{30}$$

Fig. 6(a) shows both the magnitude frequency response and the intermodulation distortion, where the peak distortion is



(g) Order 7



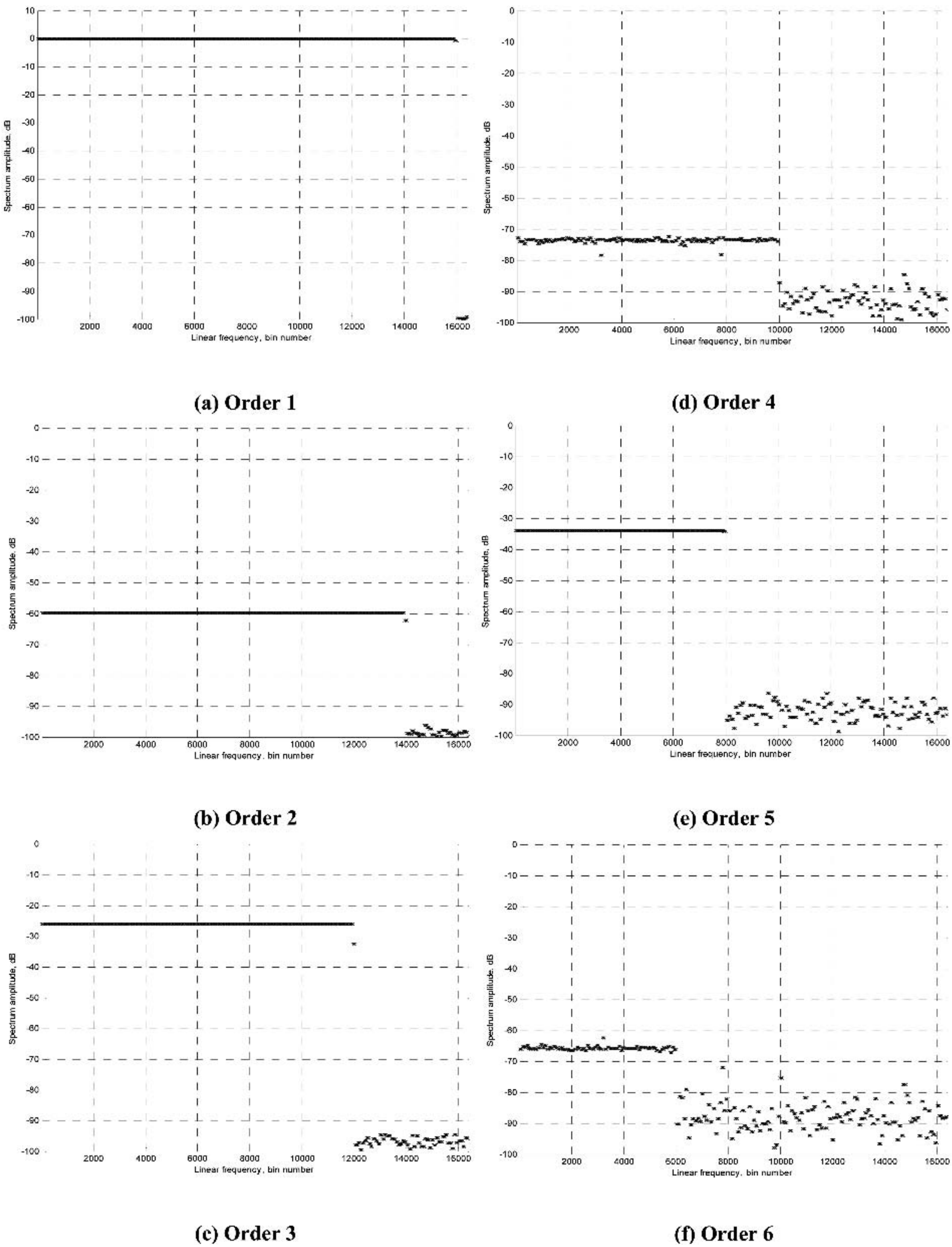
(h) Order 8

Fig. 3. *Continued*

about 40 dB below the input sequence. In this diagram the frequency response appears almost as a straight line, although with closer scrutiny Fig. 6(b) reveals the magnitude spectrum to have noiselike deviation about unity created by the noise excitation interacting with the nonlinearity.

3.3.3 Nonlinearity with Filtering

The third example employs the same nonlinearity as used in Section 3.3.2, but with the inclusion of a sixth-order Chebyshev low-pass filter with a 5-kHz bandwidth.



(a) Order 1 (d) Order 4 (b) Order 2 (e) Order 5 (c) Order 3 (f) Order 6

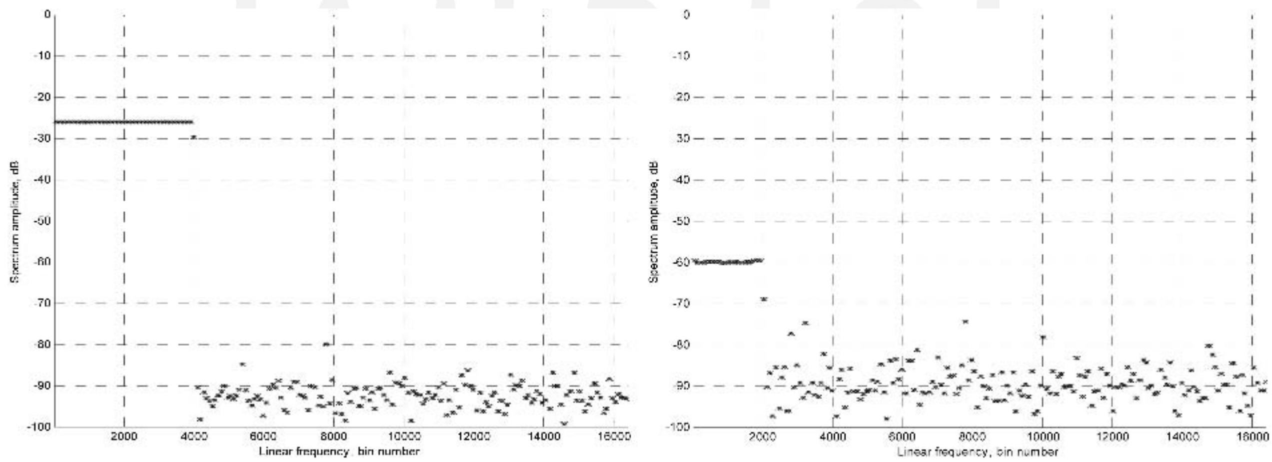
Fig. 4. Volterra frequency-domain responses for nonlinear system with memory.

Three cases were simulated: 1) filter only without nonlinearity; 2) filter located before the nonlinearity; 3) filter located after the nonlinearity.

Derived magnitude responses and distortion spectra for these three cases are shown in Figs. 7–9. The results for just the filter confirm accurate identification of the transfer function, with low distortion levels and only mild levels of noise shaping and progressive spectral corruption as the filter gain approaches the measurement noise floor. When the nonlinearity is positioned after the filter, Fig. 8 now reveals frequency-shaped distortion that follows the attenuation characteristic of the low-pass filter. Finally, Fig. 9 shows again the filter response, but here the filter predictably band-limits the broad-band distortion created by the nonlinearity.

4 SYSTEM TESTING USING MUSIC SIGNALS

This section investigates three nonlinear system examples using a periodic music excitation combined with the procedures described in Sections 1 and 3. The first is a memoryless nonlinearity, the second an MP3 codec, and the third a desktop editor sampling-rate converter. The rationale for choosing music is that certain nonlinear audio systems such as perceptual codecs produce excitation-specific distortion critical to their operational philosophy. Also, because system modeling is nonfeasible for many classes of nonlinearity, the relationship between excitation and distortion to auditory masking [18] establishes the foundation for perceptually motivated objective analysis (see Holler et al. [19], [20]). Two variations of the mea-



(g) Order 7 (h) Order 8

Fig. 4. *Continued*

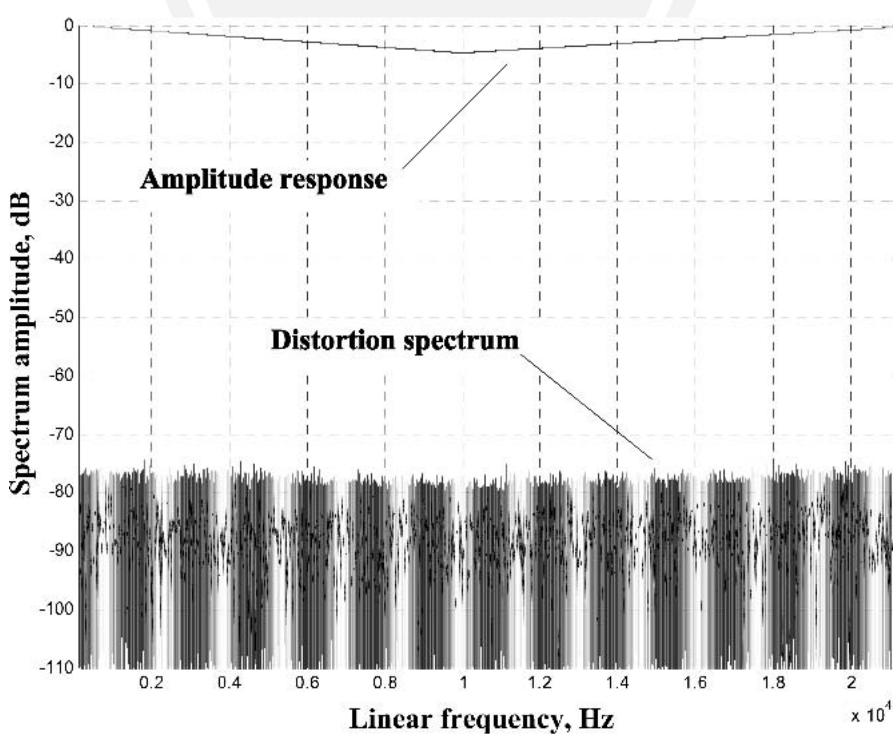
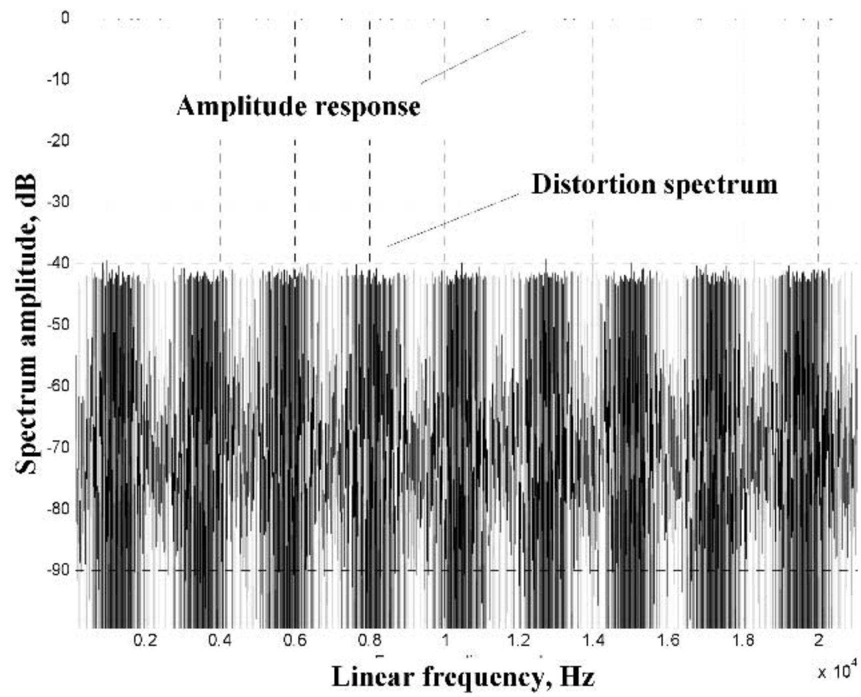
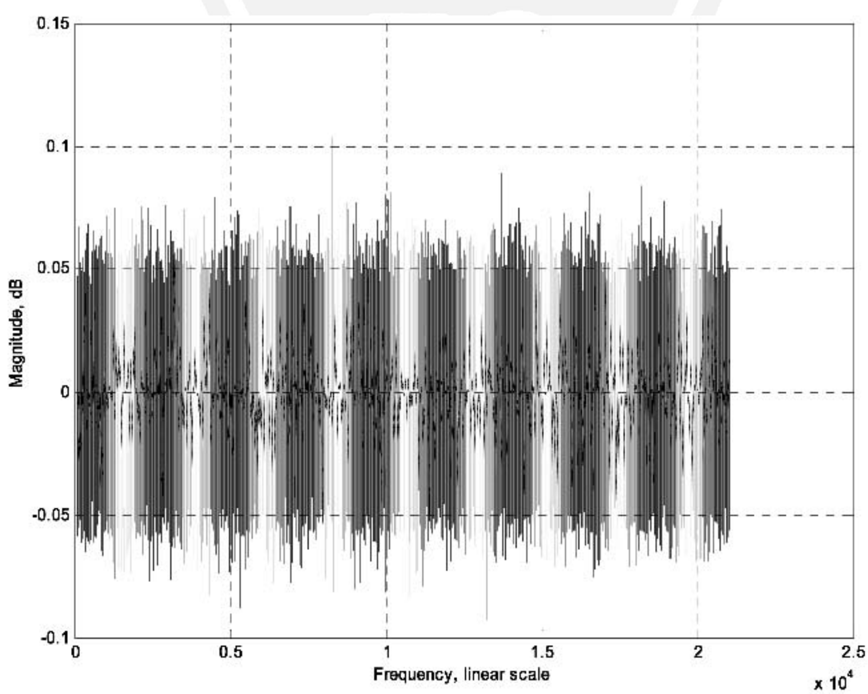


Fig. 5. Measured amplitude response and distortion spectrum of linear filter.

surement system are presented. The first, described in Section 4.1, uses a full-frame music sequence combined with comb filtering, with the spectral analysis applied full frame, a technique better suited to non-time-varying systems. In the second variation, presented in Section 4.2, the test sequence is modified so that signal excitation and distortion generation become uniquely localized in time. Also spectral analysis of both excitation and distortion is applied to overlapping data blocks of a duration of typically 25 ms. These two expedients enable the measurement system to be applied to time-varying systems and thus include, for example, perceptual codecs. The objective is to display how the short-term excitation spectrum tracks the short-term signal distortion, thus facilitating the inclusion of more sophisticated masking models to enable formal perceptually motivated analysis. In the example presented in Section 4.2, a short zero signal segment is embedded in the input in order that a corresponding null in the distortion spectrum can be observed to validate correct temporal linkage between excitation and distortion. Also, in order to gain greater insight into how the comb filters are visualized in the time domain, Appendix B.4 presents a z-domain analysis of the overall process, including signal generation, comb filter-



(a)



(b)

Fig. 6. (a) Distortion derived using memoryless nonlinearity. (b) Noiselike magnitude frequency response for memoryless nonlinearity.

ing, and data analysis, whereas Appendix B.5 describes the method of block-based Fourier analysis used to form the three-dimensional spectral temporal-frequency output displays.

4.1 Memoryless Nonlinearity Evaluated Using Music Excitation

For comparison, the same nonlinearity as used in Section 3 was tested with music [see Eq. (30)] and comb-filter processing. However, the excitation consisted now of an $N = 2^{20}$ sample sequence of music, requantized with dither[5] from 16 to 24 bit to extend measurement resolution. Fig. 10 shows both full-frame signal and distortion spectra. Interestingly here the distortion spectral envelope is revealed to be similar to that of the music signal.

[5]Dither is set at the 16-bit level, generated to 32-bit resolution.

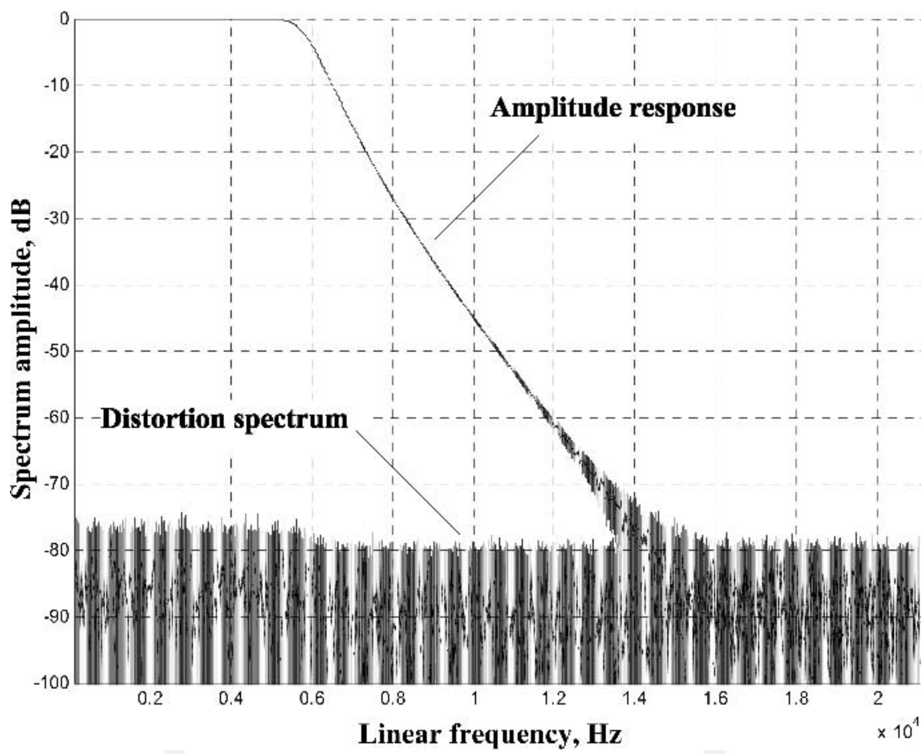


Fig. 7. Spectral results derived using full-resolution interleave procedure for low-pass filter.

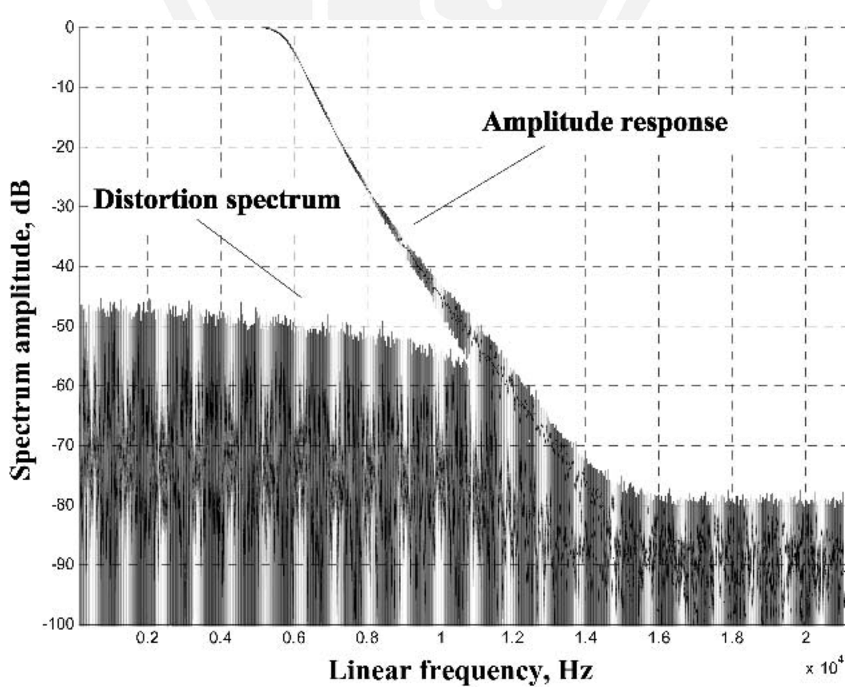


Fig. 8. Spectral results derived using full-resolution interleave procedure for low-pass filter followed by memoryless nonlinearity.

4.2 Perceptual Codec Evaluated Using Music Excitation

Two illustrative measurements were performed on an MP3 codec operating at 192 kbit/s. The first was based on spectral interleave analysis, whereas the second adapted the nonfiltered procedure of Section 1 to extract true distortion using a finely calibrated difference technique. When applying spectral analysis across a whole music frame, as in Section 4.1, although the magnitude response is a faithful average assessment, the distortion spectrum is unrealistic because with perceptually motivated coding the error spectrum undergoes dynamic modulation in an attempt to match the masking behavior of the human auditory system. However, a more representative performance evaluation can be solicited by applying short-term spectral

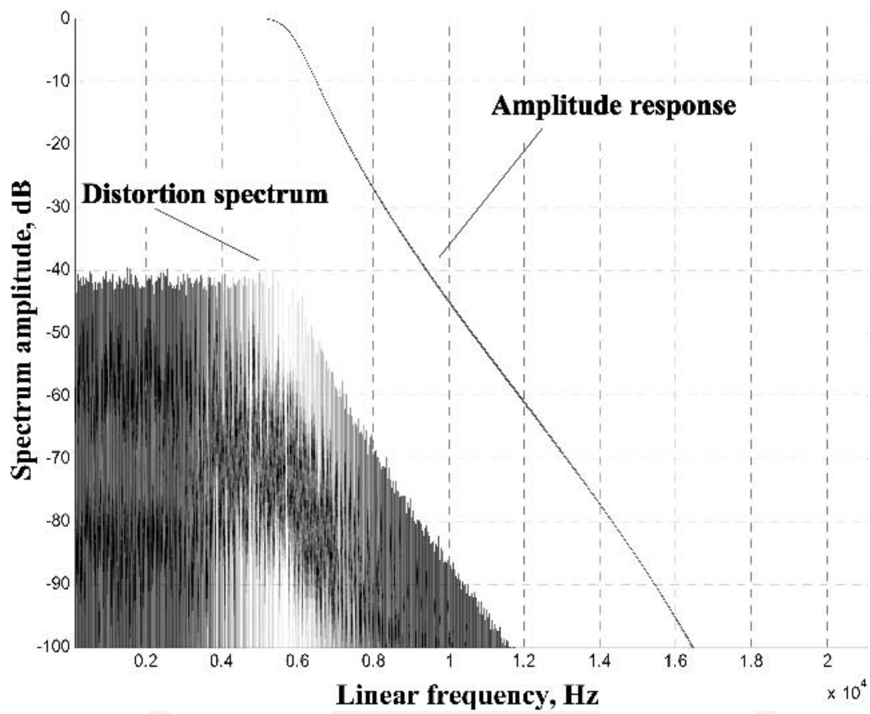


Fig. 9. Spectral results derived using full-resolution interleave procedure for low-pass filter preceded by memoryless nonlinearity.

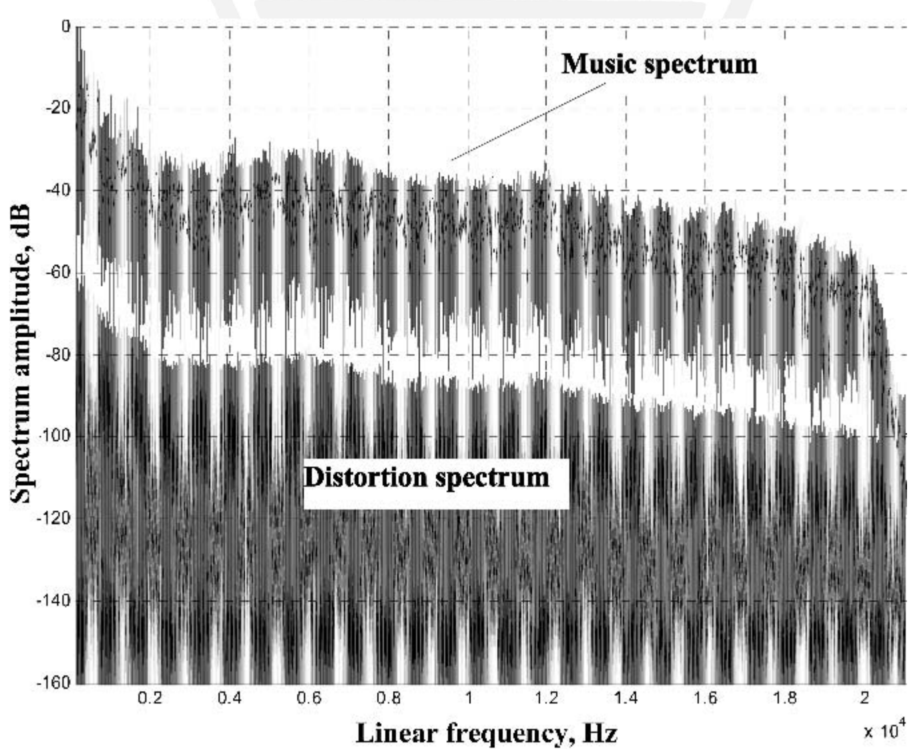


Fig. 10. Music signal and distortion spectra for memoryless nonlinearity.

analysis (see Appendix B.5) to the measurement time-domain output sequences $out_f(n)$ and $dist_f(n)$ derived from Eqs. (28) and (29).

4.2.1 MP3 Codec Example Evaluation Using Comb-Filter Analysis

For a measurement to have relevance in the context of a perceptual codec, distortion and signal must be coherent in time. In practice this condition is not met when spectral interleaving and full-length test sequences are used as comb filtering introduces circular time dispersion to both the excitation and the subsequently recovered distortion (see Appendix B.4). However, by applying a rectangular window to force frame samples in the range $N/2$ to N to zero in the music source, the following changes arise:

1) Eq. (47) forces $test_1(1{:}m/2) = test_1(1 + m/2{:}m)$ without time-aliasing distortion.

2) The codec now experiences the test sequence twice in each analysis frame, but because of the stochastic processes within a perceptual codec the distortion generated in each repeated sequence should be similar in terms of its spectral envelope but lacking coherence due to phase noise; thus $dist_1(1{:}m/2) \neq dist_1(1 + m/2{:}m)$.

3) In decoding distortion, although Eq. (50) implies a zero result if the distortion waveform were repeated precisely when the excitation is repeated, noncoherence implies that a significant fraction of the distortion is retrieved.

4) As a corollary, if Eq. (50) yields a low output incompatible with the expected level of distortion, this implies coherence and actually reveals poor randomization of coding artifacts.

By way of example, Fig. 11 shows spectral-domain results for output and distortion for an MP3 codec at 192 kbit/s. A short zero-level gap was included within the music sequence to confirm the temporal coincidence of signal and distortion. This zero signal segment is clearly resolved in both spectral displays.

4.2.2 MP3 Codec Example Evaluation Using Difference Test

To eliminate problems of time dispersion using comb filters and to allow the full vector sequence to be used in analysis, the system was adapted to enable the input–output error to be determined. The measured signal vector of length N was recovered as described in Section 1, and both excitation and measured vectors were normalized to have identical standard deviation. Circular correlation together with circular data shifting was then used to achieve precise time alignment and to correct for time delay in the codec, allowing true distortion to be calculated by subtraction.

The measured data and derived distortion are finally processed to produce a dynamic spectrum using the same block analysis as in Section 4.2.1 (see Appendix B.5). The spectral-domain result for the distortion (using the same music segment with a zero gap to facilitate comparison) is shown in Fig. 12(a) whereas Fig. 12(b) presents the corresponding plot of signal spectrum minus distortion spectrum [see Eq. (60)]. Although the details of the distortion spectra derived using the two techniques differ, a similar form is evident.

4.3 Desktop Audio Editor Sample-Rate Conversion

As an example of the use of the comb-filter-based measurement system to evaluate algorithms within a desktop audio editor, the procedure was applied to both integer and noninteger sample-rate conversion. The exploration had two stages of sample-rate conversion and converted audio data initially at 44.1 kHz in two directions, such that the output file sampling rate returned the same rate as the input file. The input frame length had 2^{20} samples and thus gave a frequency resolution of about 0.042 Hz. The results are shown in Fig. 13. It can be seen that for integer sample-rate conversion there is virtually no distortion evident whereas for the noninteger conversion, apparent high-level distortion has been generated. In fact the distortion remained low in both cases, but the algorithm introduced small block-based frequency shifting errors during conversion due to the noninteger conversion ratio. Thus signal frequency-dependent spillage into the adjacent null bands was resolved, as shown in Fig. 13(b).

5 GRAPHICAL DISPLAY OF SMALL RESPONSE ERRORS

To conclude the discussion on system measurement, this section describes a means of representing small frequency-response deviations more accurately. As an illustration, consider a discrete echo of time delay τ_{echo} and relative amplitude γ_{echo}, where the discrete frequency-domain transfer function $G(n)$ of the system is

$$G(n) = 1 + \gamma_{echo} * \exp(-i2\pi n f_0 \tau_{echo}) \tag{31}$$

with f_0 being the block repetition frequency of the excitation sequence and $n = 1{:}N$. The magnitude response shows periodic frequency variation, where the peak-to-peak response variation Dev_{dB} about the target response is

$$Dev_{dB} = 20 * \log_{10}((1 + \gamma_{echo})/(1 - \gamma_{echo})). \tag{32}$$

Table 1 expresses Dev_{dB} as a function of γ_{echo}. If presented graphically, as $\gamma_{echo} << 0$ dB, it becomes progressively more difficult to discern Dev_{dB} and therefore to extract the true character of the error, implying that fine detail may be lost and measurement data misrepresented. To improve the representation of small response deviations, an error function $E(n)$ is defined in terms of the measured transfer function $G(n)$ and the target function $T(n)$,

$$G(n) = T(n).*(1 + E(n)). \tag{33}$$

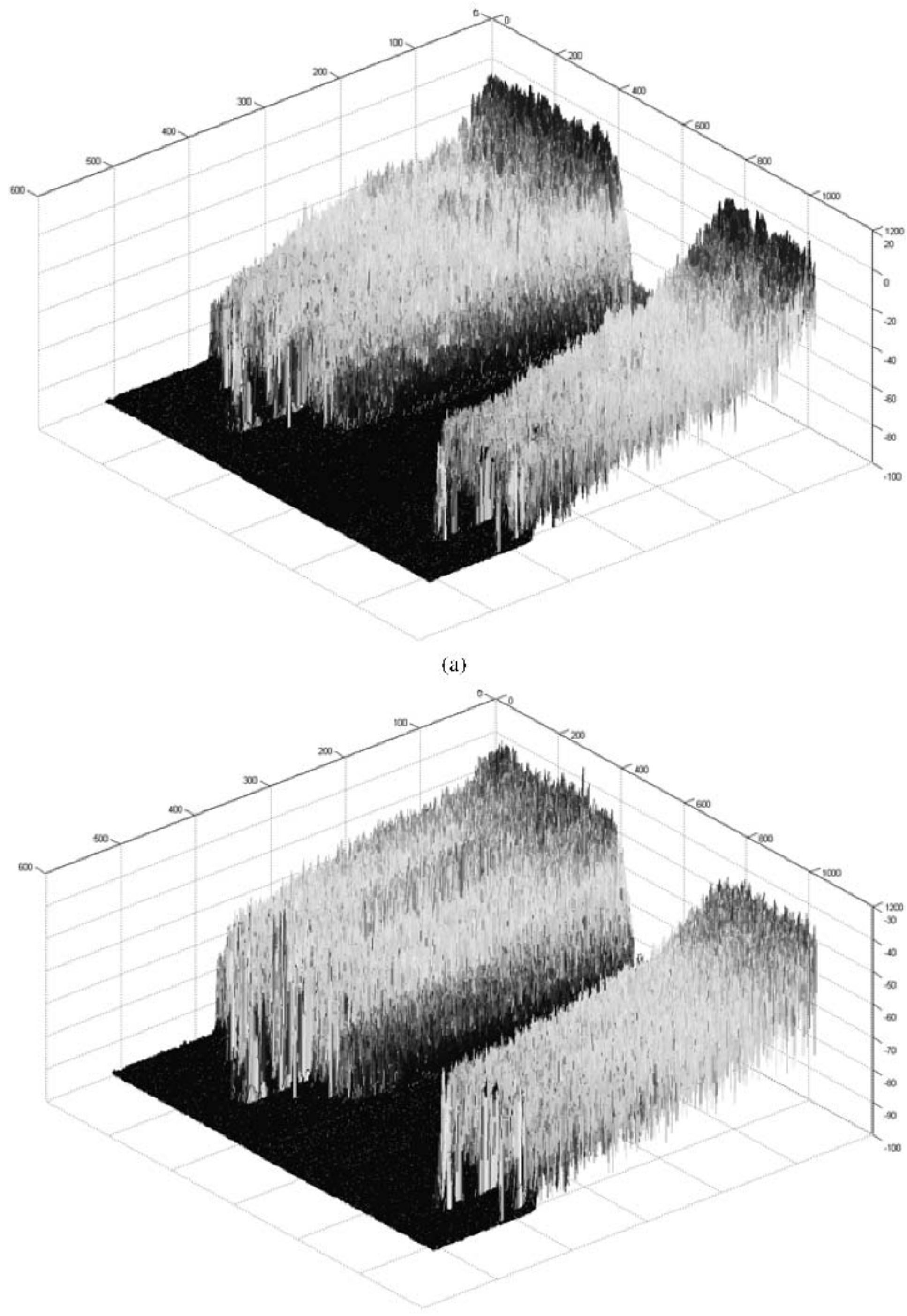
Fig. 14 shows a representation of Eq. (33), where the error function can be referred to either the input or the output of the system. In practice, to impart more performance information, both amplitude response and error function can be plotted on the same graph, where the distance between traces forms a measure of LDR. As an example, Fig. 15 shows a three-dimensional plot of the frequency response and the error response resulting from the single echo. It can be seen that although frequency response deviations become harder to discern as the echo level is reduced, the error information is retained in the error function plot.

To use the LDR display when the target system is unknown requires frequency response estimation. It is assumed that the target response, although not necessarily flat, is characterized by a smooth curve allowing, for example, spline interpolation to form a smooth curve fit to a frequency subsampled (typically a factor in the range 32 to 256) version of the measured amplitude response. To circumvent phase problems, a magnitude-based error spectrum referred either to the input or to the output may be defined in terms of the actual measured spectrum and smoothed spectrum as follows:

$$\text{magnitude error spectrum}|_{\text{dB,input}} = 20 * \log_{10}(10^{-10} + \text{abs}(\text{abs}(G(n)) - \text{abs}(T(n)))) \quad (34)$$

$$\text{magnitude error spectrum}|_{\text{dB,output}} = 20 * \log_{10}(10^{-10} + \text{abs}(\text{abs}(G(n))./\ \text{abs}(T(n)) - 1)). \quad (35)$$

As a final evaluation example, a CD player with integral upsampling was measured [21] using the test procedures described in Section 1. Fig. 16 shows both the magnitude



(a) (b)

Fig. 11. MP3 codec performance at 192 kbit/s derived using interleave filter. (a) Codec output spectrum. (b) Codec distortion spectrum.

frequency response (top trace) and the error function (lower trace) presented on a common graph, where the target response was estimated using spline interpolation. The space between the two traces defines the LDR of the CD player. Response errors result from both quantization artifacts and frequency ripple within the interpolation filters used within the DAC. Fig. 17 shows the actual input minus output spectral error, confirming that ripple close to the noise floor is resolved and also demonstrating the accuracy achieved by the measurement system. (Note that some spectral lines below 2 kHz are believed to be low-level interference and not related to the system under test.)

6 CONCLUSIONS

A PC-based measurement system has been described that exploits either pseudorandom noise or music and where measurement accuracy is bounded mainly by external converter performance. In addition to transfer function and distortion measurements, the scheme included a simplified Volterra model as a method of nonlinear modeling where, although not universal, it forms a compromise between full system identification and measurement-specific assessment techniques. The procedures were examined using a number of linear and nonlinear examples, where

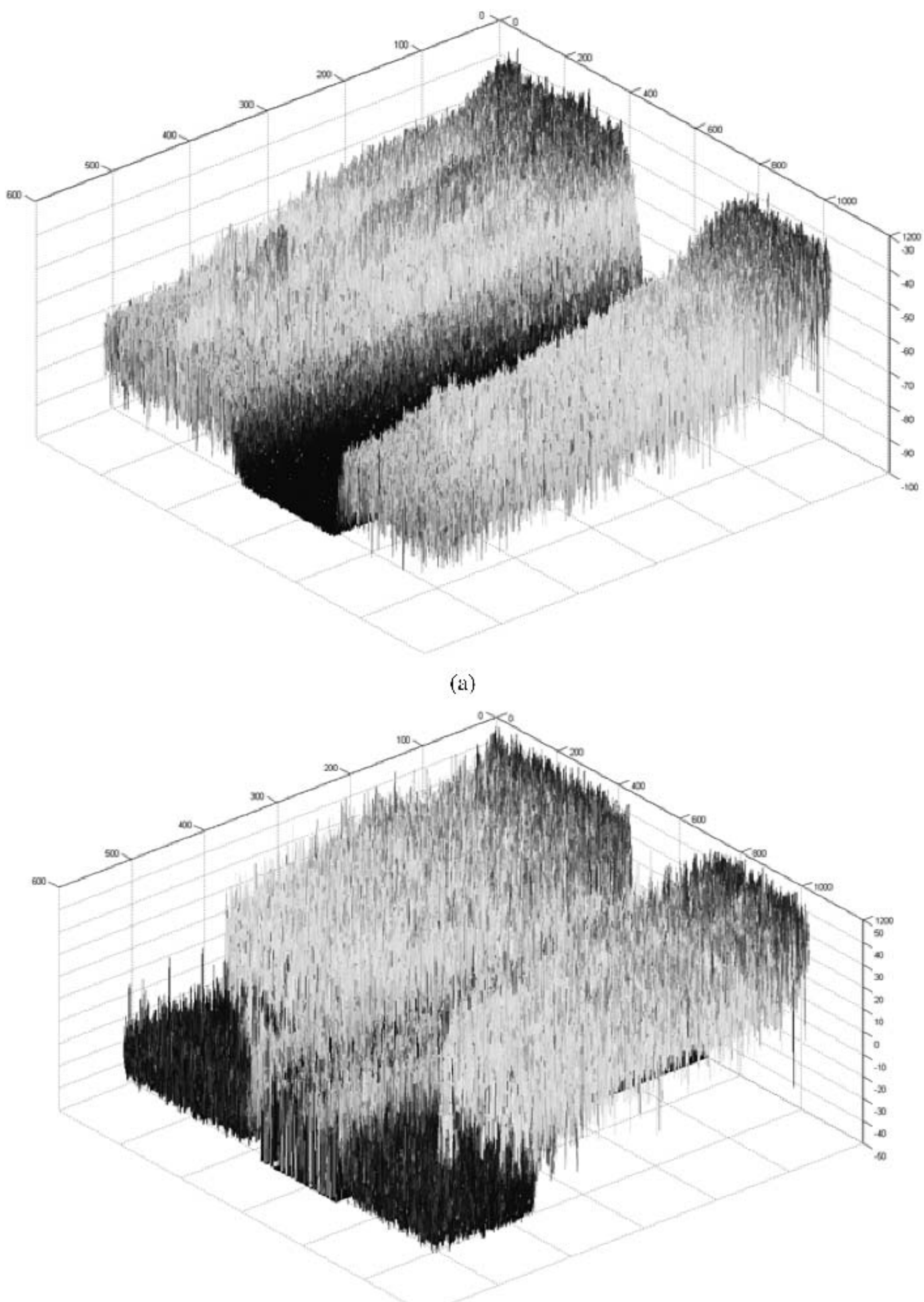


Fig. 12. MP3 codec performance at 192 kbit/s derived using true difference test. (a) Codec distortion spectrum. (b) Codec signal spectrum minus distortion spectrum.

resolution and accuracy were confirmed and insight was gained into the way the distortion was spectrally shaped according to excitation spectrum and nonlinearity.

The testing regime was extended to include comb filters in both signal generation and data analysis. This allows signal and distortion to be separated within the frequency domain, enabling estimates of frequency response and distortion to be made with a single-pass measurement. Also, by including complementary comb filters and excitation virtually full measurement resolution is achieved with minimal filtering artifacts present in the recovered output signal. The use of a single-pass test signal is important, not only to save time, but to minimize effects of gain drift that otherwise contribute to measurement error.

Methods were also reported using music sequences for system evaluation, and two example applications were presented. In particular, the opportunity to evaluate time-varying systems such as perceptual-based codecs was described, where standard block-based analysis was included to enable time-varying spectral distortion and signal-to-distortion information to be displayed.

In applying these techniques there are three principal caveats to be observed.

Caveat 1 As with MLS measurement systems, the analysis is based on circular transforms, and it is critical for the excitation sequences to have a duration that ex-

Table 1. Peak-to-peak (dB) deviation as a function of echo amplitude.

Echo (Error) Level (dB)	Peak-to-peak Frequency Response Variation Dev_{dB}
−10	5.6884
−20	1.7430
−30	0.5495
−40	0.1737
−50	0.0549
−60	0.0174
−70	0.0055
−80	0.0017
−90	5.4934e-004
−100	1.7372e-004

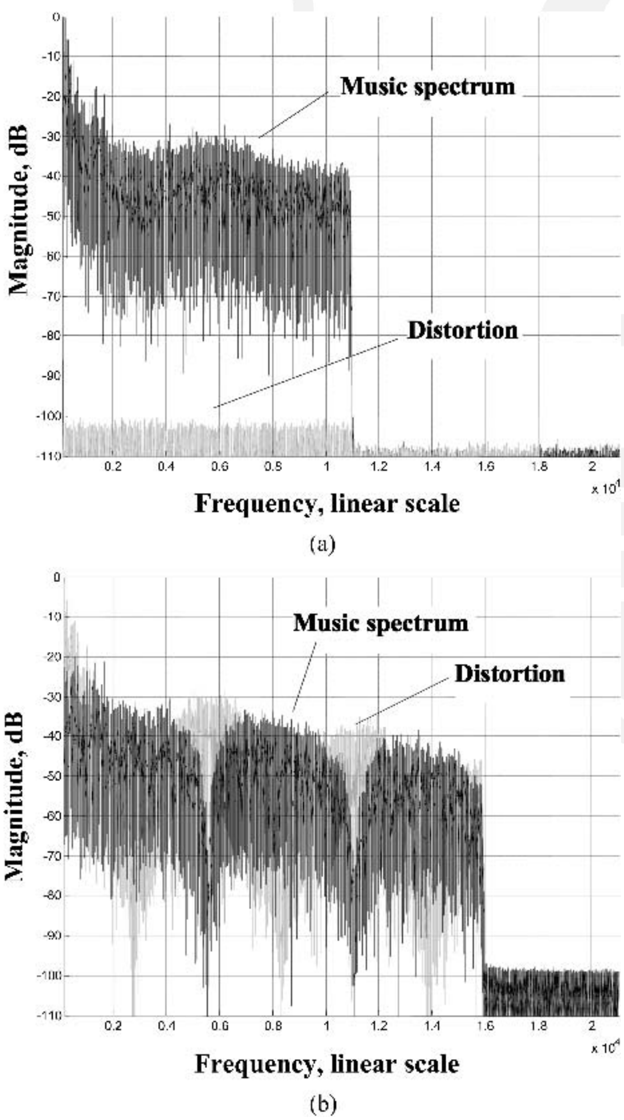


Fig. 13. Error spectra derived using music signal excitation and interleave processing for sampling-rate conversion. (a) Integer-ratio sampling-rate conversion. (b) Non-integer-ratio sampling-rate conversion.

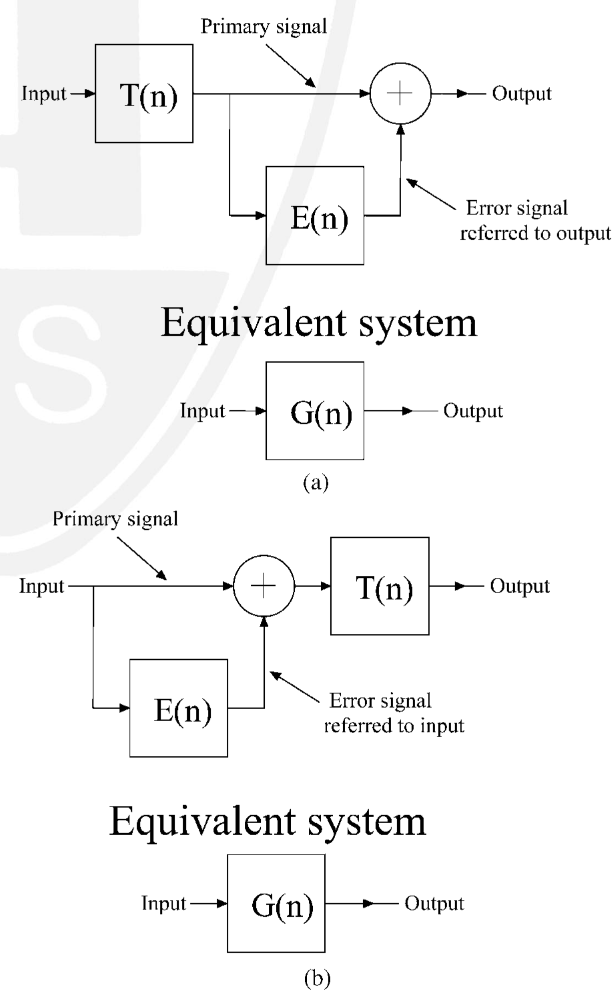


Fig. 14. Transfer function with error referred to both output and input.

ceeds the time over which the measured system's impulse response remains significant. Because of time-aliasing distortion it is not possible to reduce the excitation period and then apply window functions as the source is repetitive. However, the recorded measurements do capture the full impulse response and as such do not show windowing artifacts. Of course, if a loudspeaker system is measured, then the standard practice of windowing the derived total impulse response to eliminate reflections can be applied. However, the duration of the test sequence must exceed not only the loudspeaker impulse response but also the effect of room reflections and subsequent reverberation.

Caveat 2 A second factor is the requirement for exact sampling-rate synchronization of the test source and the ADC used to capture the measured response. This is especially critical when comb filters are incorporated as otherwise spectral spillage degrades the separation of distortion and transfer function data. All analyses and discussions presented have assumed sampling-rate synchronization. However, the need for precise frame alignment is less important as multiple sequences are output and transform circularity applies. Thus framing error just adds uncertain delay to measured impulse responses but has no effect on magnitude transfer functions.

Caveat 3 The process of comb filtering allows the separation of some of the intermodulation distortion. However, the filters viewed in the time domain introduce circular time dispersion where the excitation is effectively repeated twice and also overlaid with the existing sequence. This is not a problem when just making an input-

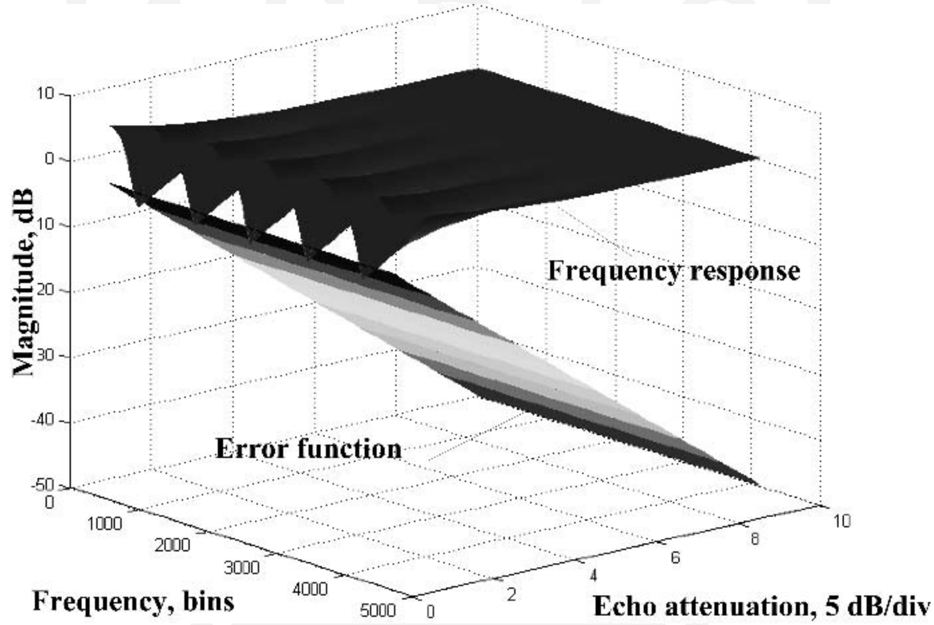


Fig. 15. LDR spectral plot for system with single echo path.

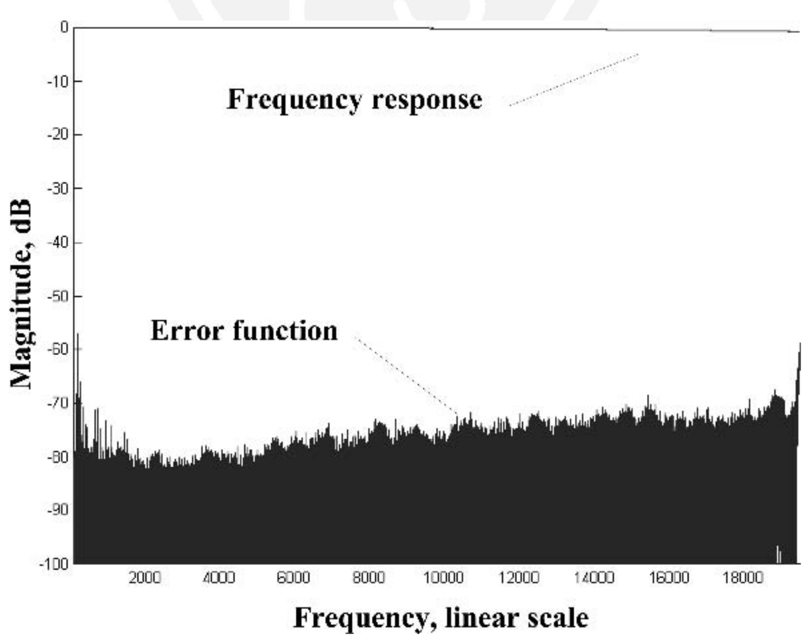


Fig. 16. Frequency response and magnitude error function for CD player.

specific distortion estimate. However, where time coherence is important, as with perceptual coding, this approach must be used with caution. This aspect was discussed in detail, and an absolute difference method was included for precise distortion analysis.

Finally, to complement the measurement schemes, two methods of enhanced data display were discussed. For linear distortion, a graph combining both frequency response and error function was proposed. In Section 5 this was shown especially suitable for cases where small response deviations occur, whereas for systems such as loudspeakers it is less suitable as the error function is relatively large. A classic three-dimensional spectral display was also included because of its relevance to perceptual codecs.

Using the system in a number of applications, the technique has proved to be an accurate and sensitive instrument to extract performance parameters. Also, it has enabled insight to be gained about the relationship between excitation and distortion spectra for a range of nonlinearities. For example, when a nonlinear system is tested using noise, then small frequency-response irregularities appear because the distortion is noiselike. When these small deviations are analyzed using the error function and LDR display described in Section 5, then the assessment of transfer function and distortion compares favorably with that derived using the comb-filter process in Section 3.

7 Acknowledgment

The author would like to offer his appreciation for the helpful feedback given by the reviewers, which has led to a number of significant enhancements of this paper.

8 References

[1] J. Borish and J. B. Angell, "An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise," *J. Audio Eng. Soc.,* vol. 31, pp. 478–488 (1983 July/Aug.).

[2] D. D. Rife and J. Vanderkooy, "Transfer-Function Measurement with Maximum-Length Sequences," *J. Audio Eng. Soc.,* vol. 37, pp. 419–444 (1989 June).

[3] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 48, p. 350 (2000 Apr.), preprint 5093.

[4] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems* (Wiley, New York, 1980).

[5] S. Boyd, Y. S. Tang, and L. O. Chua, "Measuring Volterra Kernels," *IEEE Trans. Circuits Sys.,* vol. CAS-30, pp. 571–577 (1983 Aug.).

[6] A. J. M. Kaizer, "The Modeling of the Nonlinear Response of an Electrodynamic Loudspeaker by a Volterra Series Expansion," presented at the 80th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 34, p. 388 (1986 May), preprint 2355.

[7] A. J. Berkhout, M. M. Boone, and C. Kesselman, "Acoustic Impulse Response Measurement: A New Technique," *J. Audio Eng. Soc.,* vol. 30, pp. 740–746 (1984 Oct.).

[8] R. C. Cabot, "Acoustic Applications of Cross Correlation," presented at the 64th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 27, p. 1020 (1979 Dec.), preprint 1544.

[9] J. Vanderkooy, "Aspects of MLS Measuring Systems," presented at the 93rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 40, p. 1048 (1992 Dec.), preprint 3398.

[10] E. Czerwinski, A. Voishvillo, S. Alexandrov, and A. Terekhov, "Multitone Testing of Sound System Components—Some Results and Conclusions, Part 1: History and Theory," *J. Audio Eng. Soc.,* vol. 49, pp. 1011–1048 (2001 Nov.).

[11] E. Czerwinski, A. Voishvillo, S. Alexandrov, and A. Terekhov, "Multitone Testing of Sound System Components—Some Results and Conclusions, Part 2: Model-

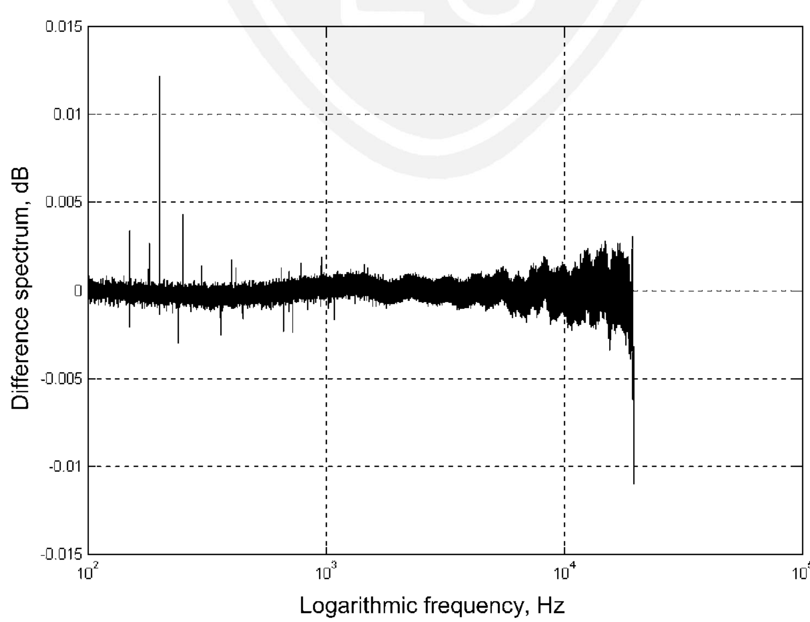


Fig. 17. Difference between estimated target response and actual response.

ing and Application," *J. Audio Eng. Soc.,* vol. 49, pp. 1181–1192 (2001 Dec.).

[12] A. I. Zayed, *Advances in Shannon's Sampling Theory* (CRC Press, Boca Raton, FL, 1993).

[13] M. O. J. Hawksford, "Digital Signal Processing Tools for Loudspeaker Evaluation and Discrete-Time Crossover Design," *J. Audio Eng. Soc.,* vol. 45, pp. 37–62 (1997 Jan./Feb.).

[14] M. J. Reed and M. O. J. Hawksford, "Identification of Discrete Volterra Series Using Maximum Length Sequences," *IEE Proc. Circuits, Devices, Syst.,* vol. 143, pp. 241–248 (1996 Oct.).

[15] M. O. J. Reed and M. O. J. Hawksford, "Efficient Implementation of the Volterra Filter," *IEE Proc. Vis., Image, Signal Process.,* vol. 147, pp. 109–114 (2000 Apr.).

[16] M. J. Reed and M. O. J. Hawksford, "Comparison of Audio System Nonlinear Performance in Volterra Space," presented at the 103rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 45, p. 1026 (1997 Nov.), preprint 4606.

[17] R. A. Belcher, "Audio Nonlinearity: A Comb Filter Method for Measuring Distortion," Rep. 2, BBC Research Dept. (1976).

[18] J. R. Stuart, "Noise: Methods for Estimating Detectability and Threshold," *J. Audio Eng. Soc.,* vol. 42, pp. 124–140 (1994 Mar.).

[19] M. P. Hollier, M. O. J. Hawksford, and D. R. Guard, "Characterization of Communications Systems Using a Speechlike Test Stimulus," *J. Audio Eng. Soc.* (*Engineering Reports*), vol. 41, pp. 1008–1021 (1993 Dec.).

[20] M. P. Hollier, M. O. J. Hawksford, and D. R. Guard, "Error Activity and Error Entropy as a Measure of Psychoacoustic Significance in the Perceptual Domain," *IEE Proc. Vis., Image, Signal Process.,* vol. 141, pp. 203–208 (1994 June).

[21] M. O. J. Hawksford, "Clockless and Bible Black," (Linn CD12 review), *Hi-Fi News* and *Record Review,* vol. 43, no. 2, pp. 66–71 (1999 Aug.).

APPENDIX A
COMMON MATLAB NOTATION AND OPERATORS

$a + \mathrm{i}b$, complex number, where $\mathrm{i} = \sqrt{-1}$

$x(n) = [x(1)\ x(2)\ \ldots\ x(r)\ \ldots\ x(N)]$, vector definition

$1{:}N = [1\ \ 2\ \ \ldots\ \ r\ \ \ldots\ \ N]$, vector with simple arithmetic progression

Element-by-Element Processing

The following three functions use the dot operator to describe an element-by-element process as distinct from conventional matrix operators. The definitions define the operations:

$x(n).{*}y(n) = [x(1)\,y(1)\quad x(2)\,y(2)\quad \ldots\quad x(r)\,y(r)\quad \ldots\quad x(N)\,y(N)]$, element-by-element multiplication of vectors $x(n)$ and $y(n)$

$x(n)./y(n) = [x(1)/y(1)\quad x(2)/y(2)\quad \ldots\quad x(r)/y(r)\quad \ldots\quad x(N)/y(N)]$, element-by-element division of vectors

$x(n).\hat{}M = [x(1)^M\ x(2)^M\quad \ldots\quad x(r)^M\quad \ldots\quad x(N)^M]$, each element in $x(n)$ is raised to the power of M

fft($x(n)$), fast Fourier transform of vector $x(n)$

ifft($x(n)$), inverse fast Fourier transform of vector $x(n)$

real($a + \mathrm{i}{*}b$) = a, real part of a complex number

abs($x(n)$) = $[|x(1)|\ |x(2)|\quad \ldots\quad |x(r)|\quad \ldots\quad x(N)|]$, magnitude value of each element

rand(1,N), vector, N random elements 0 to 1, rectangular probability distribution function

ones(1,N), vector of length N with unit elements

zeros(1,N), vector of length N with zero elements

APPENDIX B

B.1 Three-Dimensional Matrix [*X*] Inversion

Because $[X]$ is a three-dimensional complex matrix (with, for example, $8{*}8{*}2^{15}$ coefficients), inversion is performed individually for each discrete frequency in the Fourier transforms. This sequential process is relatively time consuming. However, it is undertaken only once for a given set of M noise vectors. The inversion is required for decoding by Eq. (15). Recalling that the inverse $M \times M$ matrix $[X]^{-1}$ is $[Z]$, then for frequency-domain bin x, $[Z(x)]$ follows,

$$[Z(x)] = [X(x)]^{-1}$$

$$= \mathrm{inv}\begin{bmatrix} X_{1,1}(x) & X_{1,2}(x) & \cdots & X_{1,s}(x) & \cdots & X_{1,M}(x) \\ X_{2,1}(x) & X_{2,2}(x) & \cdots & X_{2,s}(x) & \cdots & X_{2,M}(x) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ X_{r,1}(x) & X_{r,2}(x) & \cdots & X_{r,s}(x) & \cdots & X_{r,M}(x) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ X_{M,1}(x) & X_{M,2}(x) & \cdots & X_{M,s}(x) & \cdots & X_{M,M}(x) \end{bmatrix}. \tag{36}$$

The inverse described in Eq. (36) is repeated for each frequency bin x over the range 1 to N (where N is the frequency vector length) and elements are concatenated to form vectors $Z_{r,s}$ that make up matrix $[Z]$. The input-specific inverted matrix $[Z]$ is then available for decoding.

B.2 Volterra Test Sequence Formation

Initially the noise sequence generator program creates a preamble consisting of 2^{10} zero elements followed by a synchronization bit pattern that are used for frame locking following a measurement. They are defined by the vectors preamble_y and preamble_x, where

$$\text{preamble}_y = \text{zeros}(1{,}2\text{\textasciicircum}10) \tag{37}$$

$$\text{preamble}_x = [\text{preamble}_y\,[0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;1\;1\;-1\;-1\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0]]. \tag{38}$$

The M uncorrelated excitation noise vectors (each of length N) used for identification are calculated [see also Section 1, Eqs. (4) and (5)] using complex exponentials, each with individual random phase, to form a set of time-domain vectors with constant-magnitude frequency responses. The individual vectors $x_r(n)$ are derived as follows. Let

$$\text{NF}_r(n) = \exp(\text{i*angle(fft(rand}(1,N)))) \tag{39}$$

where transforming to the time domain,

$$x_r(n)|_{r=1}^{M} = \text{real(ifft(NF}_r(n))) \tag{40}$$

and forcing the mean of the vector to zero, gives the set of vectors as

$$x_r(n)|_{r=1}^{m} = x_r(n) - \text{mean}(x_r(n)). \tag{41}$$

The ensemble of M noise sequences is then amplitude normalized by the peak absolute value of the whole ensemble such that when the composite sequence is formed, signal excursion is bounded within the range −1 to 1. Also, normalization is combined with amplitude quantization and appropriate dither to set vector resolution (typically 24 bit). M subframes are then assembled using each noise vector repeated four times. Finally the test sequence gen is formed by concatenating the M subframes and the preamble,

$$\begin{aligned}\text{gen} = [&\text{preamble}_x\,[x_1(n)\,x_1(n)\,x_1(n)\,x_1(n)][x_2(n)\,x_2(n)\,x_2(n)\,x_2(n)]\\ &\cdots[x_r(n)\,x_r(n)\,x_r(n)\,x_r(n)]\\ &\cdots[x_M(n)\,x_M(n)\,x_M(n)\,x_M(n)]\,\text{preamble}_y]'.\end{aligned} \tag{42}$$

The stereo wavfile of gen with the sampling rate f_{sam} and resolution bit is realized in Matlab as

wavwrite([gen, gen], fsam,bit,′file name′).

B.3 Decoding and Analysis of Measurement Data

Following data acquisition using a sample-rate synchronized ADC, M measured data sequences designated $y_1(n)$, $y_2(n)$, . . . , $y_M(n)$ are extracted by sample counting and transformed to the frequency domain. The frequency-domain vector Y_r corresponding to excitation r is

$$Y_r\Big|_{r=1}^{M} = \text{fft}(y_r(n))/N. \tag{43}$$

The Volterra frequency-domain responses $[H]$ are determined by applying Eq. (15). Noting that $Z_{r,s}$ is the vector (r, s) of matrix $[Z]$, then the rth row of $[H]$ is calculated as

$$\begin{aligned}H_r\Big|_{r=1}^{M} = Z_{r,1}.{*}Y_1 + Z_{r,2}.{*}Y_2 + \cdots + Z_{r,3}.{*}Y_r + \cdots\\ + Z_{r,4}.{*}Y_M\end{aligned} \tag{44}$$

where both Y_r and $Z_{r,s}$ are frequency-domain vectors of length N. The set of M vectors forming H are subsequently transformed using Eq. (16) to determine the Volterra time-domain impulse-response matrix $[h]$. To correct for gain error between test sequence and redigitized measured data, $[h]$ is normalized to set the peak absolute value of the linear time-domain impulse response to unity.

B.4 *z*-Domain Description of Interleave Filtering

In this appendix time dispersion resulting from comb filtering is analyzed to give additional insight into the measurement procedure and also to inform how the system can be modified to cope with measurements of systems such as perceptual coders. The use of frequency interleaving as described in Section 3 offers a method to segregate signal and distortion and where by using two test segments with complementary interleave functions, full frequency resolution is obtainable. However, the employment of interleave filtering has consequences in the time domain in terms of both signal excitation and subsequent processing used to extract signal and distortion. In practice comb filtering introduces time dispersion, which modifies both excitation and recovery of distortion such that they are no longer time coherent, which can invalidate the results for nonstationary systems. For example, if a full-frame music signal is used, then a time-delayed version is overlaid, thus corrupting the excitation. Postprocess filtering also smears the resulting distortion such that signal and distortion are no longer linked correctly in time. For a perceptual codec evaluation this linkage is critical.

The even and odd frequency raised-cosine comb filters can be defined in the z-domain by finite impulse response filters (FIRs), where the respective filter functions $C_{\text{even}}(z)$ and $C_{\text{odd}}(z)$ are

$$C_{\text{even}}(z) = 0.25_Z^{\;0.5N} + 0.5 + 0.25z^{-0.5N} \rightarrow 0.5(1+\cos(\pi fT))$$

$$C_{\text{odd}}(z) = 0.25_Z^{\;0.5N} + 0.5 - 0.25z^{-0.5N} \rightarrow 0.5(1-\cos(\pi fT)).$$

However, circularity implies over period N that $z^{0.5N} \equiv z^{-0.5N}$, whereby

$$C_{\text{even}}(z) = 0.5(1 + z^{-0.5N}) \tag{45}$$

$$C_{\text{odd}}(z) = 0.5(1 - z^{-0.5N}). \tag{46}$$

The measurement process uses up to two sequential data sets (that is, $h = 1$, $h = 2$) using complementary comb filters, where source(z), $\text{test}_h(z)$, $\text{dist}_h(z)$, and $\text{mdata}_h(z)$ are respective samples of source sequence (noise or music), test sequence, distortion, and captured data, and $\text{decode}_{h1}(z)$, $\text{decode}_{h2}(z)$ are data decoded by the respective comb filters. All samples form elements within frames of length N.

The following z-domain system functions describe the measurement process with frame sequential, complementary comb filtering, $h = 1, 2$,

- Test sequence:

$$\mathrm{test}_1(z) = C_{\mathrm{even}}(z)\mathrm{source}(z)$$
$$\mathrm{test}_2(z) = C_{\mathrm{odd}}(z)\mathrm{source}(z)$$

- Measured response:

$$\mathrm{mdata}_1(z) = \mathrm{test}_1(z) + \mathrm{dist}_1(z)$$
$$\mathrm{mdata}_2(z) = \mathrm{test}_2(z) + \mathrm{dist}_2(z)$$

- Decoded response 1:

$$\mathrm{decode}_{11}(z) = \mathrm{mdata}_1(z)C_{\mathrm{even}}(z)$$
$$\mathrm{decode}_{21}(z) = \mathrm{mdata}_2(z)C_{\mathrm{odd}}(z)$$

- Decoded response 2:

$$\mathrm{decode}_{12}(z) = \mathrm{mdata}_1(z)C_{\mathrm{odd}}(z)$$
$$\mathrm{decode}_{22}(z) = \mathrm{mdata}_2(z)C_{\mathrm{even}}(z).$$

Rearranging and substituting $C_{\mathrm{even}}(z)$ and $C_{\mathrm{odd}}(z)$ from Eqs. (45) and (46),

$$\begin{aligned}\mathrm{decode}_{11}(z) &= 0.25(1 + z^{-0.5N})^2\mathrm{source}(z)\\ &\quad + 0.5(1 + z^{-0.5N})\mathrm{dist}_1(z)\\ \mathrm{decode}_{12}(z) &= 0.25(1 + z^{-0.5N})(1 - z^{-0.5N})\mathrm{source}(z)\\ &\quad + 0.5(1 - z^{-0.5N})\mathrm{dist}_1(z)\\ \mathrm{decode}_{21}(z) &= 0.25(1 - z^{-0.5N})^2\mathrm{source}(z)\\ &\quad + 0.5(1 - z^{-0.5N})\mathrm{dist}_2(z)\\ \mathrm{decode}_{22}(z) &= 0.25(1 - z^{-0.5N})(1 + z^{-0.5N})\mathrm{source}(z)\\ &\quad + 0.5(1 + z^{-0.5N})\mathrm{dist}_2(z).\end{aligned}$$

Observing circularity over N, where $z^{-N} \equiv 1$,

$$(1 + z^{-0.5N})^2 = 1 + 2z^{-0.5N} + z^{-N} \equiv 2(1 + z^{-0.5N})$$
$$(1 - z^{-0.5N})^2 = (1 - 2z^{-0.5N} + z^{-N}) \equiv 2(1 - z^{-0.5N})$$
$$(1 + z^{-0.5N})\,(1 - z^{-0.5N}) = (1 - z^{-N})^2 \equiv 0$$

and the equations simplify to

$$\mathrm{test}_1(z) = 0.5(1 + z^{-0.5N})\mathrm{source}(z) \tag{47}$$
$$\mathrm{test}_2(z) = 0.5(1 - z^{-0.5N})\mathrm{source}(z) \tag{48}$$
$$\begin{aligned}\mathrm{decode}_{11}(z) &= 0.5(1 + z^{-0.5N})\mathrm{source}(z)\\ &\quad + 0.5(1 + z^{-0.5N})\mathrm{dist}_1(z)\end{aligned} \tag{49}$$
$$\mathrm{decode}_{12}(z) = 0.5(1 - z^{-0.5N})\mathrm{dist}_1(z) \tag{50}$$
$$\begin{aligned}\mathrm{decode}_{21}(z) &= 0.5(1 - z^{-0.5N})\mathrm{source}(z)\\ &\quad + 0.5(1 - z^{-0.5N})\mathrm{dist}_2(z)\end{aligned} \tag{51}$$
$$\mathrm{decode}_{22}(z) = 0.5(1 + z^{-0.5N})\mathrm{dist}_2(z). \tag{52}$$

Combining the two complementary filtered sets of measurement, we make the following observations.

$$T(z) = \sum_{h=1}^{2} \mathrm{test}_h(z).$$

From Eqs. (47) and (48),

$$T(z) = \mathrm{test}_1(z) + \mathrm{test}_2(z) = \mathrm{source}(z). \tag{53}$$

$$\mathrm{DEC}_1(z) = \sum_{h=1}^{2} \mathrm{decode}_{h1}(z).$$

From Eqs. (49) and (51),

$$\begin{aligned}\mathrm{DEC}_1(z) &= \mathrm{decode}_{11}(z) + \mathrm{decode}_{21}(z)\\ &= \mathrm{source}(z) + 0.5(1 + z^{-0.5N})\mathrm{dist}_1(z)\\ &\quad + 0.5(1 - z^{-0.5N})\mathrm{dist}_2(z).\end{aligned} \tag{54}$$

$$\mathrm{DEC}_2(z) = \sum_{h=1}^{2} \mathrm{decode}_{h2}(z).$$

From Eqs. (50) and (52),

$$\begin{aligned}\mathrm{DEC}_2(z) &= \mathrm{decode}_{12}(z) + \mathrm{decode}_{22}(z)\\ &= 0.5(1 - z^{-0.5N})\mathrm{dist}_1(z) + 0.5(1 + z^{-0.5N})\mathrm{dist}_2(z).\end{aligned} \tag{55}$$

Eqs. (47)–(52) show how time dispersion affects both the test signal and the extracted distortion, where in all cases filtering superimposes a copy of the specific sequence, but with a half-frame circular delay. These results show that both the excitation and the recovered distortion exhibit time dispersion based on half-frame circular repetition. Finally, Eq. (53)–(55) describe the z-domain process applied over two sets of frames with complementary comb filtering. It is shown in Section 4 that by knowing the time-domain structure, the test sequences can be modified so that, for example, music signals remain intact (though of shorter duration), thus allowing proper analysis of perceptual codecs. This is verified by introducing a short gap in the music so that a corresponding gap in the distortion can he confirmed.

B.5 Block Analysis to Derive Spectral Envelope as a Function of Time

To display spectral distortion as a function of time, Fourier analysis is applied to a windowed segment of measured data. A raised-cosine window is used with 50% block overlap and a block length of approximately 25 ms to match the data structure common to perceptual codecs. Data are then displayed on a three-dimensional graph with respective axes of spectral amplitude, frequency, and block number (corresponding to half-block time increments of 12.5 ms). The postprocessing analysis is described for vector length N as follows.

For compatibility with the FFT, the block length is 2^W, where W is a positive integer. Thus if the measurement sampling rate is f_s Hz, we have $2^W/f_s \approx 25$ ms, where w is calculated as

$$w = 2\text{\^{}round}(\log_2(0.025\, f_s)). \tag{56}$$

The raised-cosine window function winc for samples $1{:}w$ is defined as

$$\text{winc}(1{:}w) = 0.5(1 - \cos(2\pi((1{:}w) - 0.5)/w)). \tag{57}$$

Data blocks of length w and spaced at $w/2$ sample increments are extracted sequentially from both $\text{out}_f(n)$ and $\text{dist}_f(n)$ then weighted by $\text{winc}(1{:}w)$. For each weighted block, Fourier transformation is performed and a matrix compiled to represent the spectral surface as a function of discrete frequency and block number. Hence for analysis block b, the respective transforms $\text{OUT}_{\text{bk}}(\text{b},1{:}w)$ and $\text{DIST}_{\text{bk}}(b,1{:}w)$ for signal and distortion are

$$\text{OUT}_{\text{bk}}(b,1{:}w) = \text{fft}(\text{out}_f(b{*}w/2{:}b{*}w/2 + w - 1).{*}\text{winc}(1{:}w)) \tag{58}$$

$$\text{DIST}_{\text{bk}}(b,1{:}w) = \text{fft}(\text{dist}_f(b{*}w/2{:}b{*}w/2 + w - 1).{*}\text{winc}(1{:}w)). \tag{59}$$

The difference spectrum $\text{DIFF}_{\text{bk}}(b,1{:}w)$ can be calculated on a logarithmic basis between signal and distortion spectral surfaces, and forms a measure of the dynamic signal-to-distortion ratio,

$$\text{DIFF}_{\text{bk}}(b,1{:}w) = 20\log_{10}\left(\frac{\text{abs}(\text{OUT}_{\text{bk}}(b,1{:}w)) + \alpha}{\text{abs}(\text{DIST}_{\text{bk}}(b,1{:}w)) + \alpha}\right) \tag{60}$$

where α bounds the display range. Hence surfaces can be plotted that represent the dynamic spectral behavior as a function of time for the codec output signal, distortion, and the difference between signal and distortion.

THE AUTHOR

Malcolm Hawksford received a B.Sc. degree with First Class Honors in 1968 and a Ph.D. degree in 1972, both from the University of Aston in Birmingham, UK. His Ph.D. research program was sponsored by a BBC Research Scholarship and he studied delta modulation and sigma-delta modulation (SDM) for color television applications. During this period he also invented a digital time-compression/time-multiplex technique for combining luminance and chrominance signals, a forerunner of the MAC/DMAC video system.

Dr. Hawksford is director of the Centre for Audio Research and Engineering and a professor in the Department of Electronic Systems Engineering at Essex University, Colchester, UK, where his research and teaching interests include audio engineering, electronic circuit design, and signal processing. His research encompasses both analog and digital systems, with a strong emphasis on audio systems including signal processing and loudspeaker technology. Since 1982 his research into digital crossover networks and equalization for loudspeakers has resulted in an advanced digital and active loudspeaker system being designed at Essex University. The first one was developed in 1986 for a prototype system to be demonstrated at the Canon Research Centre and was sponsored by a research contract from Canon. Much of this work has appeared in *JAES,* together with a substantial number of contributions at AES conventions. He is a recipient of the AES Publications Award for his paper, "Digital Signal Processing Tools for Loudspeaker Evaluation and Discrete-Time Crossover Design," for the best contribution by an author of any age for *JAES,* volumes 45 and 46.

Dr. Hawksford's research has encompassed oversampling and noise-shaping techniques applied to analog-to-digital and digital-to-analog conversion with special emphasis on SDM and its application to SACD technology. In addition, his research has included the linearization of PWM encoders, diffuse loudspeaker technology, array loudspeaker systems, and three-dimensional spatial audio and telepresence including scalable multichannel sound reproduction.

Dr. Hawsford is a chartered engineer and a fellow of the AES, IEE, and IOA. He is currently chair of the AES Technical Committee on High-Resolution Audio and is a founder member of the Acoustic Renaissance for Audio (ARA). He is also a technical consultant for NXT, UK and LFD Audio UK and a technical adviser for *Hi-Fi News* and *Record Review.*