Audio Engineering Society

Convention Paper 10225

Presented at the 147th Convention
2019 October 16 – 19, New York

Using Volterra series modeling techniques to classify black box audio effects

Ethan R. Hoerr and Robert C. Maher

Electrical and Computer Engineering, Montana State University, Bozeman, MT USA 59717-3780

Correspondence should be addressed to Ethan Hoerr (`ethan.hoerr@student.montana.edu`)

ABSTRACT

Digital models of various audio devices are useful for simulating audio processing effects, but developing good models of nonlinear systems can be challenging. This paper reports on the in-progress work of determining attributes of black-box audio devices using Volterra series modeling techniques. In general, modeling an audio effect requires determination of whether the system is linear or nonlinear, time-invariant or –variant, and whether it has memory. For nonlinear systems, we must determine the degree of nonlinearity of the system, and the required parameters of a suitable model. We explain our work in making educated guesses about the order of nonlinearity in a memoryless system, and then discuss the extension to nonlinear systems with memory.

1 Introduction

With the proliferation of digital signal processing in both software and embedded hardware implementations, emulating existing audio systems is a natural curiosity. There is no shortage of evidence to support this: consider the contemporary work on modeling nonlinear audio systems such as distortion pedals, guitar amplifiers, and audio limiters [1, 2, 3]. If an audio system of interest exhibits linear, time-invariant (LTI) behavior, methods such as periodic impulse excitation, maximum-length sequences, and time-delay spectrometry exist to represent the system with either its impulse response or frequency response function [4, 5, 6]. However, many interesting audio devices exhibit nonlinear and/or time-varying properties, rendering the well-known LTI techniques ineffective. In particular, this paper is concerned with nonlinear system modeling considerations.

In theory, the ability to abstract audio devices sounds convenient, but what goes into the process of determining a digital model? The process of replicating a given system depends on our prior knowledge of its inner workings. At one extreme, if we completely know the electronic circuit of a device we can create a parametric or "white box" model of the effect; on the other hand, if we know nothing about the device, we employ a variety of "black box" modeling techniques, which mandate a degree of assumptions [1]. Supposing an optimal solution exists within our chosen model framework, we seek to minimize the error between the actual and modeled system. To improve our model it is helpful to identify basic system properties such as memory, linearity, time-invariance, and causality.

Assuming time-invariance, stability, causality and no memory, by benchmarking the audio device-under-test (DUT) against a linear Volterra series model of the system, we hope to glean information regarding the system's degree of nonlinearity. With this information, we hope to take a more guided approach to determining a Volterra series model resulting in a lower overall mean-square-error compared to the DUT.

The organization of this paper is as follows. First, we provide a review of basic properties from systems theory, providing context when dealing with audio systems. Second, we present a summary of approximating audio systems with Volterra series models, and discuss considerations when modeling causal, stable, time-invariant, and memoryless systems that may be nonlinear. Then, we generate linear and nonlinear models of known systems in an attempt to establish rules of thumb in deciding degree of nonlinearity. Finally, we summarize the practical considerations and limitations of this approach and how it may apply to real-world systems.

2 Black Box Model Considerations

To aid discussion of system modeling, we first review the key system properties of memory, linearity, time-invariance, causality, and stability from [7]. We wish to contrast the conveniences of modeling linear time-invariant systems with the difficulties of modeling nonlinear systems. If a system is linear and time-invariant (LTI), it can be represented by an impulse response or frequency response function. However, we will see that this framework breaks down if linearity is violated.

2.1 Basic System Properties and their implications

Memory: In a system with memory, there exists some mechanism to retain energy or information such as a capacitor, kinetic energy, or digital memory. The instantaneous output of a system with memory may be a function of the present input as well as past or future inputs. In contrast, the output of a memoryless system is only dependent on the input at the present time. Example audio systems with memory include reverberation units or a filter, whereas a memoryless system could be a simple gain block.

Linearity: A linear system obeys the superposition principle [7]. When considering sinusoidal input signals, a linear system's output will be some linear combination or superposition of the input frequencies, whereas a nonlinear system may include nonlinear combinations of the input frequencies. A fuzz pedal and tube amplifier are known to exhibit nonlinear behavior, manifested by harmonic distortion [8].

Time-invariance: In a time-invariant system, a time-shift of the input signal results in the same time-shift of the output signal. Intuitively, the state or settings of a time-varying system vary with time. While a wah pedal at a fixed position can be viewed as time-invariant, modulating the pedal position manifests the time-varying nature of the system.

Causality: A causal system only operates on present and past inputs. All physical systems are causal as they cannot anticipate future inputs. A real-time pitch shifter is a causal system due to only operating on present and past inputs, whereas a real-time implementation of an audio time-reversal would require knowledge of future inputs that have not yet occured.

Stability: A system is stable if inputs with bounded amplitude produce outputs whose amplitude is also bounded; conversely, an unstable system may produce outputs with unbounded amplitude. In an orientation lending to feedback, a microphone connected to an amplifier and loudspeaker is example of an unstable system resulting in an unbounded growth of the output amplitude. A stable audio system could be the impulse response of a room, whose total acoustic energy decays with each wave reflection/absorption.

2.2 Audio System Assumptions

In this paper, we make the assumption that our systems are causal, stable, time-invariant, and memoryless, but may be linear or nonlinear. In the case of a nonlinear system with memory, a suitable black box framework is the Volterra series representation [9]. An example memoryless nonlinear system is that of a Taylor series expansion [8]; we demonstrate that this type of system may be modeled by the Volterra series, which is essentially a Taylor series with memory [10].

3 Volterra series modeling approach

Schetzen [11] gives an illuminating presentation of the Wiener and Volterra theories of nonlinear systems, which are both ways of relating the output of a time-invariant system $y(t)$ to an input $x(t)$. While it is possible and valid to produce a Wiener or Volterra series model given a characteristic equation for a known system, our concern here is with the empirical determination of a "black box" Volterra series model based on a cross-correlation measurement technique, known as the Lee-Schetzen method. In brief, this method exploits statistical properties of white Gaussian noise to isolate and measure Wiener kernels corresponding to each degree of nonlinearity of a system. It is important to note that the measured Wiener kernels are dependent on the variance of the identification signal. The Wiener kernels can be thought of as impulse responses of various dimensions from zero (DC term), one (linear impulse response), up to infinity. The nonlinear behavior of a system is captured by the n-dimensional impulse responses for $n \geq 2$. We then obtain the homogeneous Volterra kernels by combining all Wiener kernels of the same order. The total response of a system can be approximated by infinite order and memory of Wiener/Volterra kernels, but in practice we must truncate order and memory. The Volterra series representation does have limitations; for example, it is not suited for modeling extreme nonlinearities such as hard clipping [5, 8].

The aforementioned Lee-Schetzen method is a an important aspect in determing the Volterra kernels of a black box system, and it is still being improved upon [9]. For example, Orcioni et al. [12] apply their novel multivariance approach to determine models of tube amplifiers, demonstrating the approach's relative root-mean-square error (RRMSE) advantage when compared to the single-variance approach of the classic Lee-Schetzen method [13]. Whereas these improvements yield lower error measurements by a different approach to the implementation, we seek to improve model error by establishing guidelines for choosing model order and memory length. Our hope is that by combining both approaches, we can produce more accurate models.

3.1 Parameters when determining a Volterra series model

Once again, the systems we are concerned with here are causal, stable, time-invariant, and memoryless, with no assumption made about linearity. When applying the cross-correlation technique to determine a Volterra series model, we have several parameters to choose: length of identification signal, identification signal variance, Volterra series model order, and model memory.

Input signal length is important as the Volterra theory is closely linked to the Wiener theory of nonlinear systems. Measurement of the Wiener kernels, which the Volterra kernels are based on, relies on Wiener G-functionals that are orthogonal when the input signal has a white Gaussian distribution [11]. In practice, the longer our test signal, the closer we approach an ideal Gaussian distribution resulting in better kernel estimation.

Input signal variance results in the excitation of different levels of nonlinearity. Consider an arctangent function: for low input amplitudes, the function may be approximated as linear. As input amplitude increases, the nonlinear behavior of the function manifests as the output is no longer a homogeneous scaling of the input. Thus, the choice of input variance of the identification signal will affect the linear or nonlinear excitation of the system.

Model order refers to the highest-order Volterra kernel we wish to identify. We want to strike a balance between low orders and low model error, as computational cost increases exponentially with additional higher-order kernels [12].

Model memory refers to the energy or information storage aspect of the Volterra series model. A system with finite memory may be compared to a system with a finite impulse response. For a system with a memoryless nonlinearity, we only need one value for each kernel, as the instantaneous output is only dependent on the instantaneous input. This is akin to a finite impulse response with only one coefficient. For systems with memory, determining the memory for kernel is a necessary step to optimizing our model. As mentioned in the previous subsection, kernels may be thought of as the n-dimensional impulse response of the system. Just as with measuring a linear impulse response, capturing the total impulse response for all dimensions is important to preserving the nature of the nonlinear system.

3.2 Determining system properties given a Volterra series model

As mentioned at the beginning of this section, modeling nonlinear audio systems with the Volterra series is well-known. In this subsection we will discuss how to make educated guesses about the linearity and memory of the DUT, which influence our choices when creating a Volterra series model. As mentioned before, we assume only that the system be causal, stable, time-invariant, and possibly nonlinear. Given some known stimulus, we use the relative root mean square error metric (RRMSE) in Equation 1 to compare the responses of the DUT and our model, y_i and $\hat{y}_i$, respectively. Qualitative metrics such as human factors listening tests are also important, but are beyond the scope of this paper.

$$RRMSE = \frac{\sqrt{\sum_{i=0}^{N}(y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=0}^{N}(y_i)^2}} \quad (1)$$

We first wish to answer the question, "is the system linear?" By comparing the harmonic content of the input to and output from the DUT, total harmonic distortion (THD) is a familiar way to check whether our system adheres to superposition, as a linear system should have no harmonic distortion [8]. Our proposed approach is to create various linear Volterra series models of the reference system, and to use (1) to quantify model performance against the actual system. Because measured Volterra kernels depend on the variance of the measurement signal, we will use multiple variances to generate multiple models. We predict that as RRMSE values increase for a given model, the linear model becomes worse at approximating the nonlinear behavior of the reference device.

Then, if the system appears to be nonlinear we ask, "to what extent is the system nonlinear?" For example, if we know our DUT is a square-law operator, we should only need up to a second-order Volterra series model. With black box systems we lack this detailed knowledge, however. By creating multiple linear models of the DUT using increasing input signal variance, we can get a rough idea of whether increased variances are exciting nonlinearities in the system.

[12] gives some guidelines to determining memory extent of an audio amplifier, which has a symmetric distribution of kernel values about some peak. To summarize, the first-order kernel is first identified using a large memory value due to low computation cost. Then, the significant kernel values are kept by truncating values below a noise threshold. Higher-order kernels also have truncated memories based around the peak of the first-order kernel, to reduce computational complexity. We focus only on memoryless nonlinearities here; the extension to systems with memory remains as a future development.

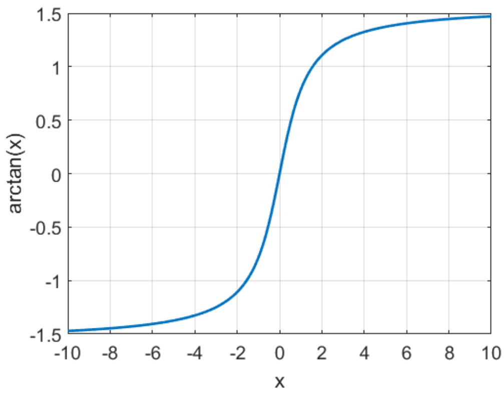

Fig. 1: Plot of arctan(x) for $x = [-10, 10]$. This sigmoidal function may be approximated linearly for small inputs x, but exhibits increasingly nonlinear behavior as x increases.

4 Experiment: Nonlinear, memoryless system

As mentioned, we seek to establish some guidelines for assessing the degree of nonlinearity of a system by conducting experiments with a known memoryless nonlinearity. Our hope is that by developing insights from a known system in a controlled context, we will have more confidence in approaching unknown, real-world black box systems.

For our memoryless, time-invariant nonlinear system we choose $y = \arctan(x)$. The response of this sigmoidal function becomes increasingly nonlinear as input amplitude increases; see Figure 1. Though our reference system is primitive compared to a real physical system, we feel this choice is justified as we see the contemporary use of the hyperbolic tangent function – also a sigmoidal function – for emulating the nonlinear behavior of distortion and guitar amplifier audio systems [1]. With a reference system decided, we next choose our experimental parameters: degree of nonlinearity of the Volterra series model, input signal duration in samples, input signal variances, and validation signal variances.

4.1 Experiment parameters

The first parameter we control is the degree of nonlinearity of the Volterra series model. We begin by creating a first-order (linear) Volterra series model of the system, then subsequently increase the order of our models by one until we reach a fourth-degree model. This is to see if adding nonlinear terms offers improvement over a purely linear approximation.

Secondly, input signal duration is important because the measurement of Wiener and thus Volterra kernels depends on identification signals with ideal white Gaussian distributions [11]. Practically speaking, we can never get a perfect white Gaussian distribution with finite-length sequences, but the approximation improves as we include more samples. We choose 10^6 samples for our input signal length, as this seems to be a good balance between computation time and nearly white Gaussian distribution of our finite data.

Our third parameter is the variance of our white Gaussian identification signal, σ_{id}^2. This parameter directly influences our measured kernel values. For instance, from [11] consider the expression for the first-order kernel $k_1(\tau)$ as a function of σ_{id}^2 and the time-average of the reference system response $y(t)$ and "n-dimensional delayed input" $y_1(t)$:

$$k_1(\tau) = \frac{1}{\sigma_{id}^2}\overline{y(t)y_1(t)}. \tag{2}$$

Higher-order kernels are also variance-dependent; see [13] for a thorough discussion of such expressions. To model the reference system under various input conditions, we will use a logarithmic range of σ_{id}^2 values from 0.01 to 5.

Fourth and finally, the variance parameter σ_{val}^2 refers to the variance of the signals used to validate each model. As noted in [9], a Volterra series kernel performs best when the input signal has a variance nearly that of the signal used to measure the kernel. RRMSE values will be computed for each combination of σ_{id}^2 and σ_{val}^2; we expect to see local minima in RRMSE where these variances are a close match. Our validation signal variance ranges logarithmically between 0.001 and 50; note that this range covers more variances than than the identification signal variance, to simulate testing models with variances beyond those expected.

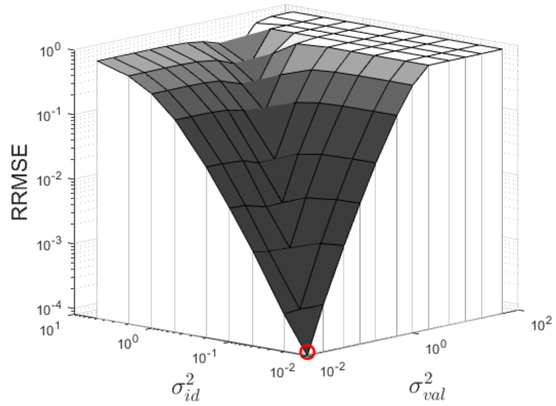

Fig. 2: RRMSE values for linear approximations of arctan(x), each obtained as a first-order memoryless Volterra kernel. A minimum RRMSE of 8.17×10^{-5} occurs where σ_{id}^2, σ_{val}^2 = 0.01.

4.2 Experiment method

A single test case consists of a choice of model order, identification signal variance σ_{id}^2, and validation signal variance σ_{val}^2. Our output from each test case is an RRMSE which compares our reference system against the generated model with the given order, σ_{id}^2, and σ_{val}^2. Within our range of experimental parameters we generate Volterra series models with the Lee-Schetzen cross-correlation method by adapting code provided in [9].

4.3 Experimental results: Linear model

In Figure 2 we see that the global minimum of our linear model occurs where input signal variance $\sigma_{id}^2 = 0.01$ evaluated with a signal variance $\sigma_{val}^2 = 0.01$. This point is marked with a circle; the RRMSE here is 8.17×10^{-5}. As we move along the axis labelled σ_{val}^2, the RRMSE increases. This is because higher variances excite the nonlinear portion of the function more than lower variances used to measure the linear Volterra kernel. Note that beyond a certain threshold, we manually clipped the plot where RRMSE ≥ 1. Errors greater than or equal to 1, according to our RRMSE equation (1), suggest a poor-fitting model. Corroborating with the findings in [9], an underside view of Figure 2 (not shown) indicates local minima where $\sigma_{id}^2 = \sigma_{val}^2$. This phenomenon suggests that a better linear approximation exists for validation input signals whose variance

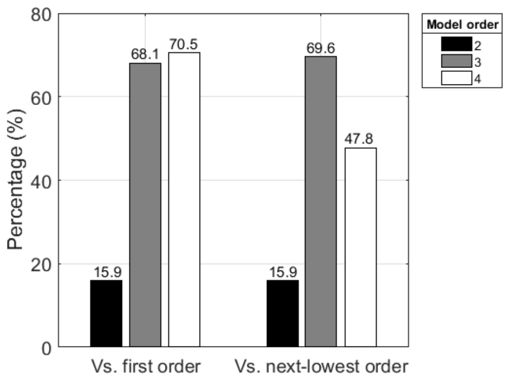

Fig. 3: Percentage of test cases with lower RRMSE compared against the linear model and next-lowest order model. Model order indicated by color code in legend.

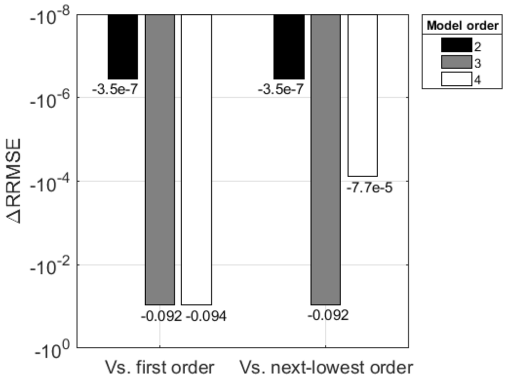

Fig. 4: Sample mean of negative ΔRRMSE test cases for each nonlinear model, compared against the linear model and next-lowest order model. Model order indicated by color code in legend.

is greater than that corresponding to the absolute minimum RRMSE.

4.4 Experimental results: Nonlinear models

With linear models established for each σ_{id}^2, we proceed to identify 2nd-, 3rd-, and 4th-order nonlinear Volterra series models. Instead of focusing on where global minima occur in each model, we are more concerned with whether the nonlinear model performs better than the linear model. As we create models that include higher-order Volterra kernels, we want a way to benchmark against our linear model.

For each test case (that is, each combination of σ_{id}^2 and σ_{val}^2), we first find the RRMSE for the various-ordered models. Then, we compare this RRMSE, $RRMSE_{Nonlinear}$, with that of the linear model, $RRMSE_{Linear}$. This results in a ΔRRMSE measure; see Equation 3. Negative ΔRRMSE values are better, indicating that the nonlinear model error is less than the linear model error. To indicate how often a nonlinear model performs better, we compute the percentage of test cases with negative ΔRRMSE's, shown in Figure 3. To get a broad idea of how much the nonlinear model improves upon the linear model, we take the sample mean of all negative ΔRRMSE's. For each nonlinear model comparison, we provide identical difference metrics against both the linear model and the next-lowest degree model's RRMSE values, respectively shown in the left and right bar groups. For example, the 3rd-degree model is compared against both the linear model and the 2nd-degree model, and so on. This helps give an idea of the relative performance of each subsequent model.

$$\Delta RRMSE = RRMSE_{Nonlinear} - RRMSE_{Linear} \quad (3)$$

5 Discussion

From the left bar groups in Figures 3-4 we can see that adding a 2nd-order term offers little RRMSE improvement over the linear model, on the order of 10^{-7}. We see that both the 3rd- and 4th-order models give a ΔRRMSE of about -0.09. Because the improvement upon the linear model offered by the 3rd- and 4th-order models is so similar, in the right bar group we show a comparison of each model order against the next-lowest order model. Although the sample mean of the ΔRRMSE for the 3rd- and 4th-order models was similar, we see that adding the 4th order term offers little improvement over the 3rd-order term, on the order of 10^{-5}. The ΔRRMSE for the 2nd-order model is the same in both bar groups because of the comparison against the linear model in both cases. The disparity between the 2nd- and 3rd-order model performance highlights the benefit of adding the 3rd-order term. Thus, if given the choice between model orders

1-4, we suspect that a 3rd-order model may best approximate the reference system. This is likely due to the Taylor series approximation for arctan(x) having only odd-order terms.

Taking a step back, the process of determining the various-ordered, memoryless Volterra series models of the reference system amounts to empirically determining a generic Taylor series approximation of the form $y(t) = \sum_{n=0}^{N} a_n x^n(t)$, where N is our highest model order, and the Taylor series coefficients a_n are a function of input variance σ_{id}^2. In a practical modeling situation, the engineer is tasked with deciding what is "good enough" for a given model. This Volterra series modeling process is equally valid if the system has memory, but we limited our focus to a memoryless nonlinearity. As mentioned earlier, including higher-order kernels may improve RRMSE at the expense of computation cost, especially when the kernels include memory.

6 Conclusion

Developing models of nonlinear audio systems can be challenging, especially if the degree of nonlinearity is unknown. We offer one approach to making educated guesses about the degree of nonlinearity of a system by comparing various measured Volterra series models against the reference system. This was done by comparing 2nd-, 3rd-, and 4th-order Volterra series models against a linear model and against each next-lowest order model to judge incremental improvements. For the memoryless, nonlinear reference system arctan(x), we saw that between orders 1-4, a 3rd-order model seemed to offer the best approximation in a relative root-mean-square error (RRMSE) sense, and that even-order nonlinear terms offered relatively low improvement. We attribute these observations to the fact that an ideal Taylor series approximation of the arctan function uses only odd-order terms.

Future extensions of the work include considering nonlinear systems with memory; that is, systems with finite impulse responses with more than one coefficient. As model order increases, the presence of memory causes computation time to increase exponentially. To address this, we might draw inspiration from [1] by approximating finite impulse responses with iteratively-determined second-order IIR filters. Also, we could combine our experimental techniques with the multivariance approach described in [9]. Although the compartive advantage of a multivariance model over a single-variance model has been shown, we theorize that the ideal choice of σ_{id}^2 for each Volterra kernel might be found experimentally to give an even better multi-variance Volterra series model. Lastly, though our error metric was quantitative, we acknowledge the importance of subjective listening tests, as a better RRMSE score does not necessarily equate to better human perception [2].

References

[1] Eichas, F. and Zölzer, U., "Gray-Box Modeling of Guitar Amplifiers," *Journal of the Audio Engineering Society*, 66(12), pp. 1006–1015, 2018, ISSN 15494950, doi:10.17743/jaes.2018.0052.

[2] Schmitz, T. and Embrechts, J.-J., "Objective and Subjective Comparison of Several Machine Learning Techniques Applied for the Real-Time Emulation of the Guitar Amplifier Nonlinear Behavior," in *Audio Engineering Society Convention 146*, 2019.

[3] A. Novak, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear System Identification Using Exponential Swept-Sine Signal," *IEEE Transactions on Instrumentation and Measurement*, 59(8), pp. 2220–2229, 2010, ISSN 0018-9456, doi:10.1109/TIM.2009.2031836.

[4] Rife, D. D. and Vanderkooy, J., "Transfer-Function Measurement Using Maximum-Length Sequences," in *Audio Engineering Society Convention 83*, 1987.

[5] Dunn, C. and Hawksford, M. J., "Distortion Immunity of MLS-Derived Impulse Response Measurements," *J. Audio Eng. Soc*, 41(5), pp. 314–335, 1993.

[6] Novak, A., Rund, F., and Honzik, P., "Impulse Response Measurements using MLS Technique on Nonsynchronous Devices," *J. Audio Eng. Soc*, 64(12), pp. 978–987, 2016.

[7] Oppenheim, A. V., Willsky, A. S., and Nawab, S. H., *Signals & systems*, Prentice Hall, Upper Saddle River, N.J., 2nd edition, 1997, ISBN 978-0-13-814757-0 978-0-13-651175-5, oCLC: 925123716.

[8] Zolzer, U., *DAFX : Digital Audio Effects, Second Edition*, John Wiley & Sons, Ltd, Chichester, 2011, ISBN 978-0-470-66599-2 0-470-66599-8.

[9] Orcioni, S., "Improving the approximation ability of Volterra series identified with a cross-correlation method," *Nonlinear Dynamics*, 78(4), pp. 2861–2869, 2014, ISSN 0924-090X, 1573-269X, doi:10.1007/s11071-014-1631-7.

[10] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proceedings of the IEEE*, 69(12), pp. 1557–1573, 1981, ISSN 0018-9219, doi:10.1109/PROC.1981.12201.

[11] Schetzen, M., *The Volterra and Wiener theories of nonlinear systems*, Krieger, Malabar (FL.), 2006, ISBN 1-57524-283-4 978-1-57524-283-5.

[12] Orcioni, S., Terenzi, A., Cecchi, S., Piazza, F., and Carini, A., "Identification of Volterra Models of Tube Audio Devices using Multiple-Variance Method," *Journal of the Audio Engineering Society*, 66(10), pp. 823–838, 2018, ISSN 15494950, doi:10.17743/jaes.2018.0046.

[13] Pirani, M., Orcioni, S., and Turchetti, C., "Diagonal Kernel Point Estimation of nth-Order Discrete Volterra-Wiener Systems," *EURASIP Journal on Advances in Signal Processing*, 2004(12), p. 145457, 2004, ISSN 1687-6180, doi:10.1155/S1110865704403011.