BLOCK-ORIENTED MODELING OF DISTORTION AUDIO EFFECTS USING ITERATIVE MINIMIZATION

Felix Eichas, Stephan Möller, Udo Zölzer

Department of Signal Processing and Communications,
Helmut-Schmidt-Universität
Hamburg, Germany
`felix.eichas@hsu-hh.de`

ABSTRACT

Virtual analog modeling is the process of digitally recreating an analog device. This study focuses on analog distortion pedals for guitarists, which are categorized as stompboxes, because the musician turns them on and off by stepping on the switch. While some of the current digital models of distortion effects are circuit-based, this study uses a signal-based approach to identify the device under test (DUT). An algorithm to identify any distortion effect pedal in any given setting by input-output (I/O) measurements is proposed. A parametric block-oriented `Wiener-Hammerstein` model for distortion effects and the corresponding iterative error minimization procedure are introduced. The algorithm is implemented in `Matlab` and uses the `Levenberg-Marquardt` minimization procedure with boundaries for the parameters.

1. INTRODUCTION

Since the first distortion stompbox had been introduced in the 1960s, these effects became very popular amongst guitarists. Some are willing to pay horrendous prices for original vintage effect pedals, others have a huge collection of distortion effects. With the aid of system identification these devices can be digitally reproduced by virtual analog modeling, providing all advantages of digital systems. Identifying analog distortion circuits and building circuit based models to capture their characteristics has widely been done in the context of virtual analog modeling.

In [1–5] circuit based approaches were used to model distortion effects. Nodal analysis is used to derive a state-space-system describing the original circuit. The state-space-system is extended to be able to handle nonlinear circuit elements. However, complete knowledge of the circuit-schematics and all characteristics of the nonlinear elements are required for this method to be applied. If the circuit-schematic of a certain device is not accessible, expensive reverse-engineering and high quality measurements would be needed to derive a digital model. Therefore a simple technique, based on input-output (I/O) measurements would be desirable to get a quick snapshot of the DUT's characteristics. An approach based on I/O measurements was already used in [6–8]. The authors use a modified swept-sine technique, originally introduced by [9], to identify an overdrive effect pedal, which is described by a block-oriented Hammerstein model. Unfortunately, the results from [6–8] could not be reproduced accurately enough from the information given in the paper for a detailed comparison.

To the authors knowledge, there does not exist an objective measure which describes the perceptual correlation between digital model and reference system output. Most of the current objec-

tive metrics to evaluate audio content, like PEAQ [10], were designed to rate the sound degradations of low bit-rate audio codecs. This work does not focus on finding an acceptable error metric but designing a proper parametric model and the corresponding identification procedure for modeling of distortion effects. The optimization is based on iterative error minimization between the parametric, block-oriented Wiener-Hammerstein model and the DUT.

In 2008 Kemper introduced a patent describing his model and identification routine for identifying nonlinear guitar amplifiers. He uses a block-oriented Wiener-Hammerstein model to emulate the characteristics of an analog guitar amplifier by analyzing the statistical distribution of pitches and volumes of the identification signal. The filters and the nonlinearity are identified by an identification procedure of his own design, analyzing small and high signal levels separately [11].

This paper is structured as follows. The model is described in Section 2. Section 3 explains the identification process. The results are described in Section 4 and Section 5 concludes this paper.

2. THE MODEL

The basic idea behind the model used in this study is to have a parametric model, which is flexible enough to adapt to many distortion effects but still simple enough to be computationally efficient. The structure of a distortion effect can be described by a Wiener-Hammerstein model. This model consists of linear-time-invariant (LTI) blocks and a nonlinear block. The blocks are ordered in series where the nonlinear block is lined by two LTI blocks. The LTI blocks are filters, which are shown in Fig. 1 as H_1
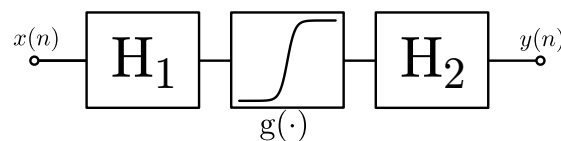
Figure 1: *Block diagram of a Wiener-Hammerstein model.*

and H_2, and the nonlinear block is basically a mapping function, mapping the level of the input signal to an output level, according to the nonlinear function $g(\cdot)$, which simulates the nonlinear behavior of the distortion effect. $x(n)$ denotes the input and $y(n)$ the output signal.

Block-oriented Wiener-Hammerstein models are successfully used in commercial products due to their flexibility and expandability. Fractal Audio Systems calls a Wiener-Hammerstein sys-
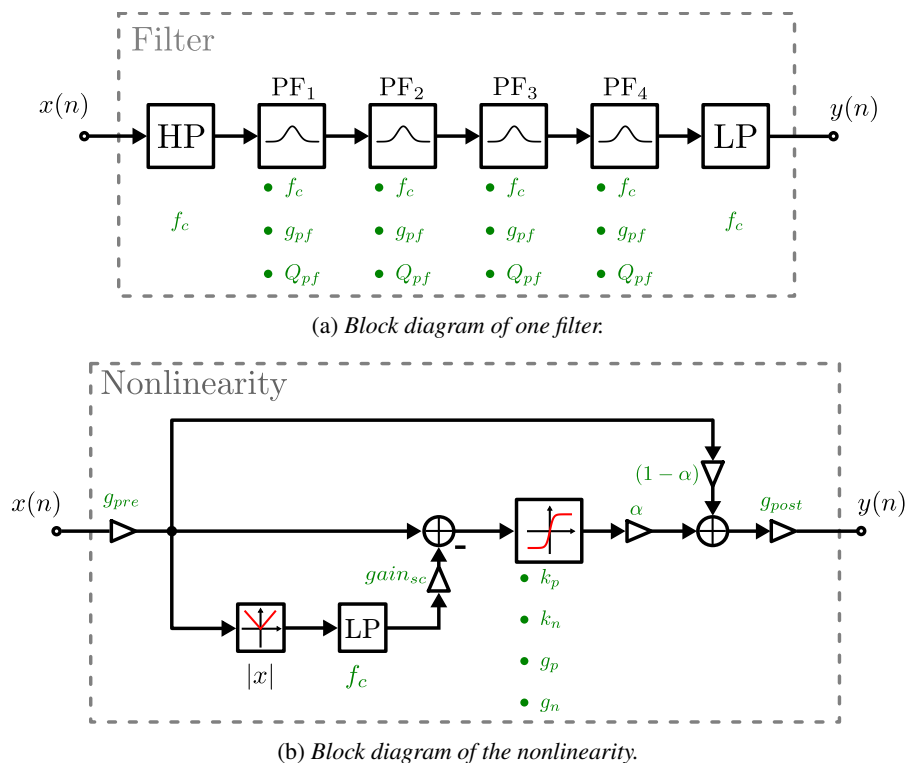
(a) *Block diagram of one filter.*

(b) *Block diagram of the nonlinearity.*

Figure 2: *Diagrams for the single blocks of the Wiener-Hammerstein model.*

tem consisting of two filters and a distorting nonlinearity the "fundamental paradigm of electric guitar tone" [12]. They extend this model with more linear and nonlinear blocks to include frequency responses of speakers and model the different nonlinear stages (preamp, power amp) of an analog guitar amplifier.

2.1. LTI Blocks

The filters of the Wiener-Hammerstein model were designed to be flexible but still stay stable in the identification process. Hence, the parameters which will be varied during optimization are not the coefficients of a Direct Form II filter structure, because the identification procedure will not converge if the filter coefficients yield an unstable filter. Instead, they are expressed as the parameters of a simple band-limited equalizer. Figure 2 (a) shows the structure of one LTI block of the Wiener-Hammerstein model. The input signal $x(n)$ is processed by a high-pass filter, a series of four peak filters and finally by a low-pass filter which yields the output signal $y(n)$. The adjustable parameters of the filters are expressed in terms of cutoff frequency f_c for the low-pass and high-pass filters. The peak filters can also be modified in terms of gain and Q-factor. All filters are second order IIR filters and their coefficients are computed according to [13].

All parameters are aligned in a parameter vector
$\mathbf{p}_{lti} = [f_{c,hp}, f_{c,pf1}, g_{pf,1}, Q_{pf1}, \ldots, f_{c,lp}]^T$.

2.2. Nonlinear Block

The nonlinear block of the Wiener-Hammerstein model is shown in Fig. 2 (b). The input signal $x(n)$ is multiplied with a pre-gain

g_{pre} and then fed into the lower side-chain which acts as a very simple envelope detector. The absolute value of the signal is computed, low-pass filtered and multiplied with the side-chain gain $gain_{sc}$. This signal is then subtracted from the direct path after the pre-gain multiplication. Subsequently it is fed into the actual mapping function, which is controlled by four parameters. Thereafter the signal can be mixed with the pre-amplified input signal via the parameter α. Finally the signal is multiplied by the post-gain g_{post} to yield the output signal $y(n)$. All parameters for the nonlinear block are aligned in the parameter vector $p_{nl} = [g_{pre}, f_c, g_{sc}, k_p, k_n, g_p, g_n, \alpha, g_{post}]^T$.

The side-chain envelope detector was added to the nonlinear block because the used mapping function is memoryless. But this is not adequate to model the time-variant behavior of e.g. vacuum tube distortion circuits. The bias-point of a vacuum tube amplification stage changes slightly, depending on the signal history, which leads to a DC-offset of the output, amongst other effects. This is emulated by subtracting the lower side-chain envelope signal from the direct path, before the mapping function is applied [14].

2.2.1. Mapping Function

The flexibility of the nonlinear block is based on the mapping function. If we would chose a mapping function based on one hyperbolic tangent, it would not be very flexible. In some cases a moderate slope around the zero-crossing of the mapping function and, at the same time, a sharp transition into the saturated region are required, because it resembles the behavior of analog circuit components. A simpler mapping function could not model that behavior accurately. Therefore, we propose an alternate mapping function,

consisting of three hyperbolic tangents which are concatenated at the location denoted by k_p for the positive part of input levels and at k_n for the negative part. The tanh functions above or respectively below k_p and k_n are modified so that they have the same slope as the middle part of the function at the connection points, shown in Eq. 1.

$$m(x) = \begin{cases} \tanh(k_p) - \left[\frac{\tanh(k_p)^2 - 1}{g_p}\tanh(g_p x - k_p)\right] & \text{if } x > k_p \\ \tanh(x) & \text{if } -k_n \leq x \leq k_p \\ -\tanh(k_n) - \left[\frac{\tanh(k_n)^2 - 1}{g_n}\tanh(g_n x + k_n)\right] & \text{if } x < -k_n \end{cases}$$
$$(1)$$

The parameters g_p and g_n control the smoothness of the transition between saturated region and linear center part. For high values of g_p and g_n the transition is very sharp, for low values it is smooth and for very small values and correctly chosen values of k_x and k_p it behaves like a linear function.
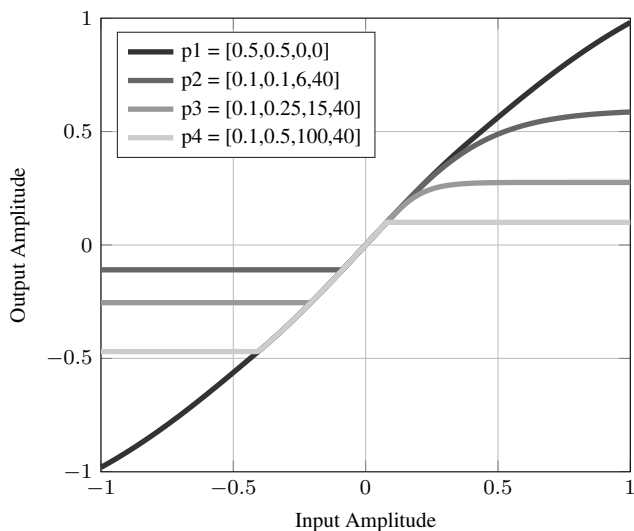

Figure 3: *Modified tanh function with the parameters* $p = [k_p, k_n, g_p, g_n]$. *Gain values for* g_p *and* g_n *are in dB.*

Figure 3 illustrates the modified tanh function. The darkest curve for parameter set $p1$ shows a nearly linear mapping function with relatively high values for $k_n > 0.5$ and $k_p > 0.5$ and low values for $g_n < 1\,\text{dB}$ and $g_p < 1\,\text{dB}$. For positive input amplitudes $p2$, $p3$ and $p4$ have steadily increasing gain values g_p and the same connection point $k_p = 0.1$. This changes the shape of the nonlinear mapping function. The gain g_n was kept constant at $g_n = 40$ dB for negative input amplitudes, while the connection parameter k_n was shifted for $p2$, $p3$ and $p4$. Positive and negative sections of each curve could be interchanged by simply changing the corresponding gain and connection parameters.

3. IDENTIFICATION

The concept of iterative error minimization is shown in Fig. 4. The same input signal is sent through the digital model and the reference system, $\mathbf{y}(n)$ denotes the desired output from the DUT, while $\hat{\mathbf{y}}(\mathbf{p}, n)$ denotes the model output, which is not only dependent on the input samples, but also on a set of parameters $\mathbf{p}$, which were introduced in Sec. 2. If the model is nonlinear for at least one
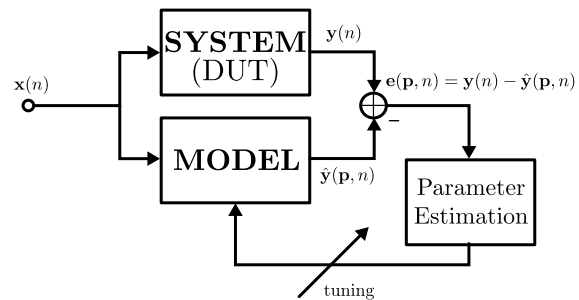

Figure 4: *Block diagram of iterative error minimization.*

parameter, it has to be identified iteratively [15]. The error signal $\mathbf{e}(\mathbf{p}, n) = \mathbf{y}(n) - \hat{\mathbf{y}}(\mathbf{p}, n)$ is calculated by subtracting the model output from the reference output. The parameter estimation algorithm calculates a new set of parameters, which are applied to the model in order to minimize the error between digital model and analog system according to a cost-function C, in our case least-squares, $C(\mathbf{p}) = \sum_{n=1}^{N} \mathbf{e}(\mathbf{p}, n)^2$. Where N is the length of the input signal in samples. The parameter estimation method used in this work is the Levenberg-Marquardt algorithm. This algorithm combines the advantages of gradient-descent and Gauss-Newton method [16, 17].

Before the identification procedure can be started, the reference signals need to be recorded. For this purpose a high quality audio interface (RME Fireface UC ™) was used, which is controlled via `Matlab`. The DUT is placed between output and input of the audio interface. Before the actual input signals are send through the DUT, the interface is calibrated by sending a test signal $x(n) = \sin\left(2\pi\frac{f_0}{f_s}n\right)$ with an amplitude of 0 dBFS, while the DUT is in bypass mode. The fundamental frequency is $f_0 = 1\,\text{kHz}$. The output gain of the interface was adjusted, so that an amplitude of 1 corresponds to 1 V at the output of the audio interface.

3.1. Nonlinear Identification

There are two different input signals for the identification of the linear and nonlinear parts of the Wiener-Hammerstein model. The input signal,

$$x_{nl}(n) = a(n) \cdot \sum_{i=1}^{M} \sin\left(2\pi \cdot \left[\mathbf{f_i} \cdot \sin\left(2\pi\frac{\mathbf{f_{mod,i}}}{f_s}n\right)\right]\right), \quad (2)$$

for the identification of the nonlinear block is created by summing several sine waves with different fundamental frequencies. Where $a(n)$ is a scaling function, creating a logarithmically rising amplitude from -60 dBFS to 0 dBFS to emphasize lower signal levels. $\mathbf{f} = (50, 100, \ldots, 900, 1000)$ Hz is a vector containing all the desired frequencies that roughly cover the range of notes that can be played on a guitar and $\mathbf{f_{mod,i}}$ is a vector containing the modulation frequencies, which range from $f_{mod,min} = 1$ Hz to $f_{mod,max} = 10$ Hz and are used to avoid destructive and constructive interference which affect the envelope of the signal. M is the amount of sine waves which are summed and then normalized to achieve a maximum amplitude of 1.

When the parameters of a nonlinear model are optimized by an iterative error minimization algorithm, like the Levenberg-Marquardt

method, it is very important to have an appropriate set of initial parameters. The minimization algorithm might converge into a local minimum if the wrong set of initial parameters is chosen. Thus, every possible combination of the parameters g_{pre}, k_p, k_n, g_p and g_n is tested on a coarse grid by calculating the sum of squares $C(\mathbf{p})$ for each combination. The set with the lowest sum of squares is used as the initial parameter set for the identification process.

Because the parameters g_{pre}, k_p, k_n, g_p and g_n have the most influence on the envelope of the input signal, this procedure helps finding a starting point, which is most likely to converge into the global minimum of the cost function. During the iterative error minimization, only the parameters of the parameter vector $\mathbf{p}_{nl}$ are optimized. The parameters of the two LTI blocks are fixed and can not be changed during this identification step. The parameters in the $\mathbf{p}_{lti}$ vectors are set to yield a neutral filter characteristic in the audible frequency range for both filters. After the optimization $\mathbf{p}_{nl}$ is saved for further use.

3.2. Filter Identification

The input signal for the parameter optimization of the LTI blocks is white noise with signal levels below -50 dBFS, because we assume, that for low signal levels the nonlinear part of the reference system operates in its linear region and $\tanh(x) \approx x$ for $|x| \ll 1$ is also true for the nonlinear block of the digital model.

In this case optimization of time domain error signals is challenging because the signal can look different in time domain, due to deviating phase characteristics of simulation and reference system, but is still perceived as similar for the human ear. For this reason the output signals $\mathbf{y}(n)$ and $\hat{\mathbf{y}}(\mathbf{p}, n)$ need special treatment before the actual minimization procedure can be started.

First the saved parameters from the identification of the nonlinear block are loaded and used during filter parameter optimization. This helps identifying both filters of the model. If the filter's parameters would be identified before the nonlinear parameters in the first optimization step, characteristics of the first filter could be optimized onto the second filter and vice versa even though the overall frequency response is retained.

The time domain output sequence for the white noise identification input is recorded. The power spectral density (PSD) of the output is computed by calculating a 16384-point discrete Fourier transform (DFT) with a hop size of 4096 samples. All calculated spectra are averaged and multiplied by its complex conjugate to yield the PSD. But the frequency resolution of a PSD or DFT respectively does not correspond to the perceptual resolution of the human ear. For this reason the semi-tone spectrum is calculated from the PSD by averaging the frequency bins corresponding to one note. The first note is A0 in the sub contra octave which has a frequency of $f_{0,A0} = 27.5$ Hz, which is one note below the lowest note on a standard tuning 5-string bass. This is done for reference and digital model respectively before the error signal is calculated.

The initial values for the identification procedure are chosen in such a way that the filter is flat and the cutoff frequencies of high- and low-pass filters are set to $f_{c,HP} = 10$ Hz and $f_{c,LP} = 18$ kHz. After the filter parameters are adapted, they are stored for further use.

3.3. Overall Identification

In this final step the stored parameter vectors for both LTI blocks and the nonlinear block are loaded and used as the initial parameter

set in the final parameter vector

$$\mathbf{p} = \begin{pmatrix} \mathbf{p}_{lti1} \\ \mathbf{p}_{lti2} \\ \mathbf{p}_{nl} \end{pmatrix}. \qquad (3)$$

The Levenberg-Marquardt algorithm is started and now all parameters of the model can be modified to refine the results from the previous optimization runs. The input signal for this step is the same as for the optimization of the nonlinear parameters, described in Subsec. 3.1. The identification is not done solely in time domain. This approach helps finding an initial parameter set for the error minimization which is likely to converge into the global minimum of the cost function. The spectrogram (8192-point DFT, 4096 hop size) of the output of the analog system and the digital simulation is calculated and then vectorized as the minimization algorithm is not able to process error signals which are not in vectorial form. By optimizing over the spectrogram error signal the necessary information for adapting the filter and nonlinear parameters is included.

4. RESULTS

As stated in Subsec. 3.2 it is challenging to find an objective error measure for the perceived differences between to audio signals. For this reason the error between simulation and reference is shown in time-domain as well as frequency-domain.

4.1. Time-Domain Error

A possible objective error measure would be the time domain error,

$$e_{y\hat{y}} = \frac{\sum_{n=1}^{N}(y(n) - \hat{y}(\mathbf{p}, n))^2}{N}, \qquad (4)$$

where N is the overall length in samples of the output signals. The error becomes zero if the signals are completely the same. But in certain cases, the time domain error is quite high but the perceived (subjective) difference is hard to hear. In general however, the time domain error gives us an estimate about how close the model can recreate the DUT. Nevertheless this is no reliable metric to characterize the perceptual difference between two audio signals. Figure 5 shows the comparison between time domain signals of DUT and the identified model. The DUT was a *Hughes & Kettner* - Tube Factor, which is basically a tube-preamp in stompbox format with a 12AX7 vacuum tube. Figure 5 (a) shows the response of digital model and reference system to an exponentially decaying sine input with 440 Hz. The maximum amplitude of the input was 1. The digital model reproduces the analog system quite well, except from the transient part at the beginning of the signal. For the recorded electrical guitar signal, depicted in Fig. 5 (b), the simulation follows the measured curve closely but there are still some differences for certain frequency components, which may be a result of the simple nonlinear block, where one nonlinear function is used for all frequency components of the input signal. Another way of determining how well the identification worked, is by comparing the transfer functions of both, reference and simulation.

4.2. Time and Frequency-Domain Error

Figure 6 shows the frequency response of a *Jim Dunlop* - Fuzz Face fuzz pedal in comparison to the frequency response of its
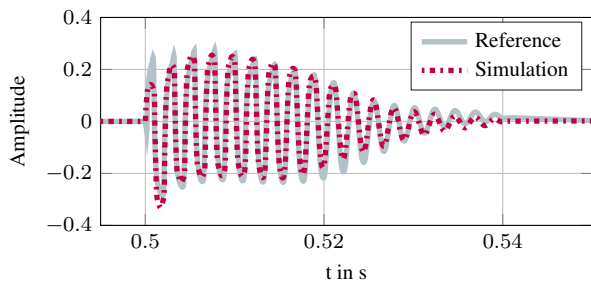
(a) 40 ms 440 Hz *sin with exponentially decaying amplitude.*
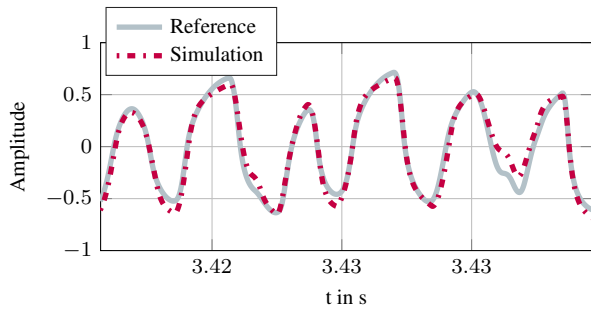
(b) *Excerpt from recorded guitar signal.*

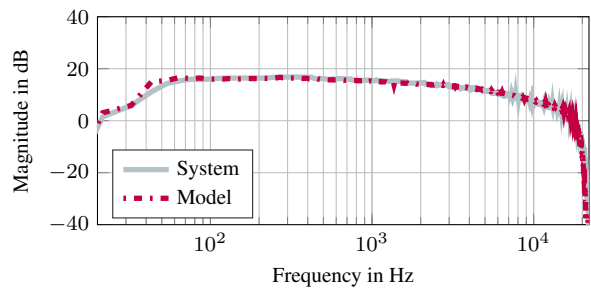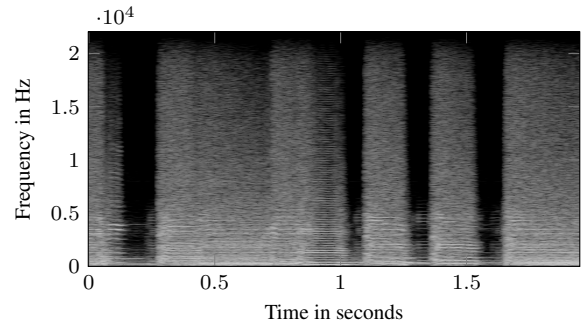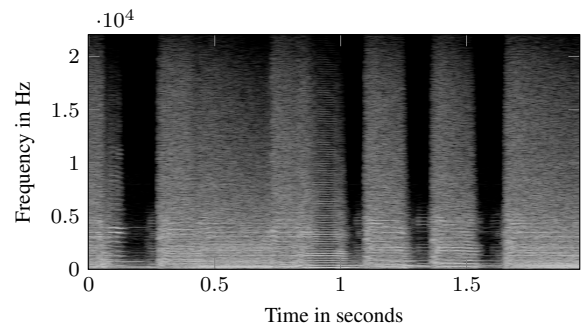Figure 5: *Comparison of the time-domain signals for the Hughes & Kettner Tube Factor and the identified model.*

Figure 6: *Comparison of the frequency response of analog reference and digital model for a Jim Dunlop - Fuzz Face.*

(a) Reference

(b) Simulation

(c) Error

Figure 7: *Spectrograms of reference system and digital model output and the error spectrogram.*

digital model. The model deviates from the reference system by less than 1 dB, but becomes more inaccurate for low frequencies (below 60 Hz) and for frequencies above 18 kHz.
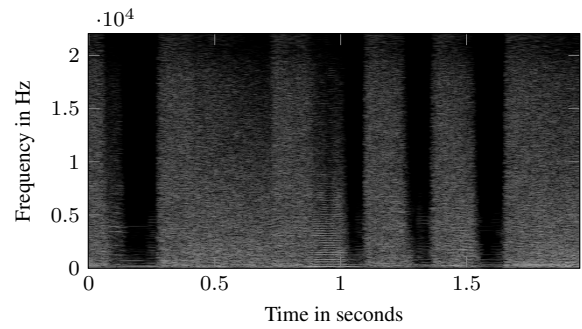
The spectrograms of digital model and analog system output are shown in Fig. 7. The input to both reference system and digital model was a recording of an electrical guitar playing fast high chords, as customary for funk music. The DUT was a *Hughes & Kettner* - Tube Factor. The harmonics generated by the DUT and its digital representation resemble each other well. The error spectrogram, shown in Fig. 7 (c), shows the difference of the absolute values of the previous spectrograms. The overall error energy is much lower than the signal energy of reference system and digital model.

4.3. Aliasing

The sampling frequency was set to $f_s = 48$ kHz for identification and runtime operations. The upper plot of Fig. 8 shows the frequency response of the digital model to a 1500 Hz sine wave. The aliasing, caused by the nonlinear block of the Wiener-Harmmerstein model is clearly visible due to the distinct peaks between the main peaks for fundamental frequency and the harmonics. For this reason resampling was introduced. The output signal of the first LTI block is upsampled by resampling factor L, then processed by the nonlinear block and finally the output of the nonlinear block is downsampled by factor L. The lower plot of Fig. 8 shows the response of the digital model to a 1500 Hz sine wave with resampling factor $L = 8$. The effects of aliasing are now nearly diminished.
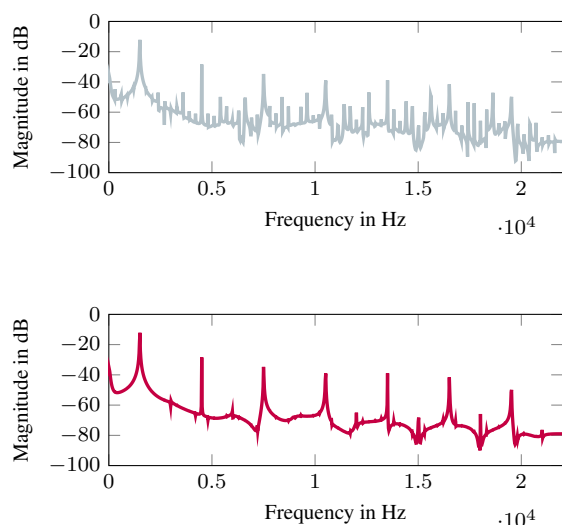
Figure 8: *Response of the digital model to a* 1500 *Hz sine. Above: without oversampling. Below: with* 8 *times oversampling.*

4.4. Auditory Impression

Although no formal listening test was conducted, the subjective auditory impression of the proposed model is quite satisfying. In some cases the difference between simulation and reference output is still audible but only for a trained listener. Different input signals as well as digital model and analog system outputs can be found online. Please visit [18] for listening examples.

5. CONCLUSIONS

This work proposed a method to identify and model nonlinear analog distortion effects. LTI filter blocks and a nonlinear block of a Wiener-Hammerstein model, are introduced. The identification routine is described and the model is able to emulate any distortion pedal in a given setting. For many effects the results from the model are nearly indistinguishable from the analog device itself. But this method still has several drawbacks, which should be addressed in the future. First the search for the initial parameter set is still carried out on a coarse grid, because the computational effort rises drastically if the grid resolution or the amount of tested parameters increases. This may cause the identification algorithm to converge into a local minimum instead of the global minimum. Furthermore it is essential to study more perceptually motivated error metrics, e.g. PEAQ, to find a comparable and reliable error metric.

6. REFERENCES

[1] D. Yeh, J.S. Abel, and J.O. Smith, "Automated physical modeling of nonlinear audio circuits for real-time audio effects: Part 1 - theoretical development," in *IEEE Trans. Audio, Speech, and Language Process.*, May 2010, vol. 18, pp. 203–206.

[2] D. Yeh and J.O. Smith, "Simulating guitar distortion circuits using wave digital and nonlinear state-space formulations,"

in *Proc. Digital Audio Effects (DAFx-08)*, Espoo, Finland, Sept. 1-4, 2008, pp. 19–26.

[3] J. Macak, *Real-time Digital Simulation of Guitar Amplifiers as Audio Effects*, Ph.D. thesis, Brno University of Technology, 2011.

[4] K. Dempwolf, *Modellierung analoger Gitarrenverstärker mit digitaler Signalverarbeitung*, Ph.D. thesis, Helmut-Schmidt-Universität, 2012.

[5] M. Holters and U. Zölzer, "Physical modelling of a wah-wah effect pedal as a case study for application of the nodal dk method to circuits with variable parts," in *Proc. Digital Audio Effects (DAFx-11)*, Paris, France, Sept. 19-23, 2011, pp. 31–35.

[6] A Novak, L. Simon, P. Lotton, and J. Gilbert, "Chebyshev model and synchronized swept sine method in nonlinear audio effect modeling," in *Proc. Digital Audio Effects (DAFx-10)*, Graz, Austria, Sept. 6-10, 2010.

[7] A. Novak, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential swept-sine signal," *Instrumentation and Measurement, IEEE Trans. on*, vol. 59, no. 8, pp. 2220–2229, 2010.

[8] R. Cauduro Dias de Paiva, J. Pakarinen, and V. Välimäki, "Reduced-complexity modeling of high-order nonlinear audio systems using swept-sine and principal component analysis," in *Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio*, Mar 2012.

[9] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Paris, France, 2000.

[10] International Telecommunication Union, "Bs.1387: Method for objective measurements of perceived audio quality," url, Available online at `http://www.itu.int/rec/R-REC-BS.1387` − accessed June 1st 2015.

[11] C. Kemper, "Musical instrument with acoustic transducer," US Patent: US20080134867 A1, July 2008.

[12] Fractal Audio Systems, "Multipoint Iterative Matching and Impedance Correction Tevhnology (MIMIC™)," Tech. Rep., Fractal Audio Systems, April 2013.

[13] U. Zölzer, *DAFx - Digital Audio Effects*, John Wiley and Sons, 2011.

[14] J. Pakarinen and D. T. Yeh, "A review of digital techniques for modeling vacuum-tube guitar amplifiers," *Computer Music Journal*, vol. 33, no. 2, pp. 85–100, 2009.

[15] T. Strutz, *Data fitting and uncertainty: a practical introduction to weighted least squares and beyond*, Vieweg and Teubner, 2010.

[16] K. Levenberg, "A method for the solution of certain problems in least squares," *Quarterly of applied mathematics*, vol. 2, pp. 164–168, 1944.

[17] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial & Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[18] "Listening Examples," Website, Available at `http://www2.hsu-hh.de/ant/webbox/audio/eichas/dafx15/audio_examples_local_version.html` − accessed June 4th 2015.