School of Physics and Astronomy

Modelling Non-Linear Behaviour in Valve Amplifiers
MPhys Project Preparation Review Report

Ewan Hemingway

1st November 2010

Signature: Date:

Supervisor: Dr. J. Kemp

1 Overview

The vast power of digital signal processing techniques have long been sought after, but it is only with fairly recent developments in computing technology that their use in real-time systems has become a reality. One active area of research, both commercially and scientifically, is in studying the behaviour of traditional valve powered guitar amplifiers. Since their invention in the early 1900s, vacuum tube amplifiers have remained an integral part of music technology and culture, despite the emergence of many cheaper, more reliable methods of amplification. In fact many "spurious" non-linear effects caused by the technology have become to be revered by musicians; the slightly ambiguous characteristic of "warmth" is also often favourably referred to. One could argue that the signal distortion characteristics of valve amps have been the catalyst for the birth of whole genres of music.

Valve amplifiers are expensive business however, with top-of-the-range amp heads costing upwards of £1000 at the time of writing [1]. They are delicate pieces of equipment, and must be transported with care. This motivates the need for a digital alternative - with the continuing advances in solid state hardware coupled with decreasing production costs, the capabilities of digital signal processing are constantly expanding, allowing for real-time performance. Multiple models can then be run on same hardware, allowing musicians to instantly switch between different classic amp sounds.

This review should demonstrate the range of current techniques in amplifier modelling, several of which are already in commercial use. These tend to be based on one of two paradigms. Black box modelling (see Sec. 2.1) tests a system with a known input signal, records the output and from this, determine the filter or filters representing the system. This methodology doesn't necessarily require any knowledge of the inner workings of the system, although it can often be difficult to generalise/parametrise the results.

The alternative approach is the White-Box method (see Sec. 2.2), which uses prior knowledge of the system to create some physical model, be it from analysing circuit components, or applying waveshapers.

For a complete simulation of the guitar setup, we must also consider the various stages in the signal path. Typically, the setup includes at least [2]:

1. **preamplifier** raises the signal level to line level, which is required to drive the power amplifier)
2. **tone stack** provides basic equalisation and tone control
3. **power amplifier** amplifies the signal to the point where the speaker can be driven

In this project, we are specifically interested in the vacuum tube which can be used in both the preamplifier and power amplifier stages. Note that the speaker itself also has a separate response, which under heavy load may exhibit non-linear behaviour, and although the methods outlined here may model this, the focus of the project remains on the tube stage of the process.

2 Literature Survey

Before continuing, it is important to characterise the different types of system encountered. First we must define the linear system. For a given input x and output $y(x)$, a linear system must be both additive and homogeneous:

$$\text{linear - } \begin{cases} y(x_1 + x_2) = y(x_1) + y(x_2) & \text{- additivity} \\ y(ax) = ay(x) & \text{- homogeneity} \end{cases}$$

Visually, as suggested by the name, a input-output plot of a linear system is just a straight line through the origin. Conversely a non-linear system doesn't have these restrictions. We find that part of what gives the vacuum tube amplifier its character is this non-linearity.

The other important characteristic is whether the system exhibits time-invariance or not. If the output is not an explicit function of t, then the system is said to be time-invariant or memoryless. The simplest class of system from an analytical point of view is the linear, time-invariant or LTI system, which we will treat first. Unfortunately, the vacuum tube amplifier is, in general, neither linear nor time-invariant.

It is also worth taking a moment to discuss the difference between continuous and discrete representations. In the digital audio domain, a continuous signal $x(t)$ is treated instead as a set of discrete samples $x(n)$, where n is the sample index. The sampling rate determines the number of samples per unit time. For the most part, this review deals with continuous variables although these representations are generally trivial to discretise for digital implementation.

With these categorisations in place, we may examine the first class of modelling: the Black Box approach.

2.1 Black Box Approaches

The Black Box approach is a completely general methodology, but it has many practical applications in acoustics. Often a system is very complex, be it an acoustic space or the circuitry of a vintage effects unit, and building a full physical model/simulation from scratch can be very difficult. In these cases, by carefully choosing a specific input signal and recording the output, we can start to model the system, even without any knowledge of its internal workings (Fig. 1).

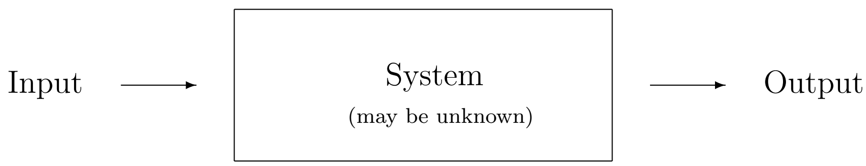


Figure 1: Schematic of a Black Box system

Here, we first treat the simplest case, the LTI system; from there we can build on this foundation to add increasing levels of complexity, adding non-linear and time-varying behaviour.

2.1.1 The LTI System

It is well accepted and understood that a linear, time-invariant system, can be fully represented by its impulse response, $h(t)$ (see [3] for example). If $h(t)$ can be found, then one can exactly recreate the output of the system by convolving it with any input signal $x(t)$:

$$y(x(t)) = h(t) \otimes x(t)$$

It is relatively simple to find the impulse response in this case. Input $x(t)$ is explicitly known, and output $y(t)$ can easily be digitally recorded. So by application of Convolution Theorem, which states that convolution in the time domain is completely equivalent to multiplication in the frequency domain [4], we get:

$$h(t) = \text{IFT}\left\{\frac{\text{FT}\{y\}}{\text{FT}\{x\}}\right\}$$

Some care must be given to the choice of input signal $x(n)$: we must avoid division by zero in the above calculation. This can be avoided by making sure all frequencies (in this digital representation) are present at some point, be this with a swept sine signal or carefully chosen noise.

2.1.2 Non-linear Time-Invariant

The next level of complexity now considers the non-linear, but memory-less system. Although is is possible to loosely class valve amplifiers under this title, in general memory effects must be accounted for thorough simulation. The methods here are generally more applicable to systems such as analogue tape recorders, though the introduction of non-linearities is an important step in understanding the valve amplifier.

Dynamic Convolution In his 2000 paper [5], Kemp suggests a novel method for dealing with these non-linearities. The main issue is that contrary to the linear case, each amplitude level can have a different impulse response. In theory you would require an infinite number of impulses to fully represent the system; in practice the dynamic convolution method only samples a discrete finite set of measurements. Kemp suggests that $M = 128$ (the number of impulses) should be sufficient for thorough simulation, although as processing power increases, this number could rise. Once the set of impulses h_1 to h_M have been normalised, we can define a selector function to choose which impulse is appropriate for the input signal level:

$$S(x(n)) = 1 + \frac{|x(n)|}{F_{sv}/M}$$

where F_{sv} is the full scale value, i.e. the maximum possible digital value at any point in time. The vertical resolution of digital audio is specified by the bit depth, so at any point in time, 16-bit audio would take one of 2^{16} values[1].

With this selector, we can modify the existing discrete convolution relation (1) to selectively choose a specific impulse (2):

$$y(n) = \sum_{i=0}^{L-1} x(n-i)h(i) \tag{1}$$

$$\downarrow$$

$$y(n) = \sum_{i=0}^{L-1} x(n-i)h_{S(x(n))}(i) \tag{2}$$

If M is as large as F_{sv} this then the simulation is as complete as can be possible digitally, but clearly this is an unrealistically large number of impulses to sample - instead approximately $M = 100$ samples are taken. However to apply these smoothly, some sort of interpolation is required, otherwise we would hear a discontinuous switch between different impulses (think of a stepped function). For simplicity, the author suggests a linear interpolation which applies a combination of two successive impulse responses and weighs their contribution appropriately. The author admits other interpolation methods are possible, and as non-linearities are at the heart of the problem, it would seem sensible to try higher order polynomial interpolations, albeit at higher computational cost.

Kemp's algorithm has attracted much commercial interest, with the software patented under a partnership with DSP hardware maker Focusrite [6]. Improvements in quality can be achieved by taking responses at higher resolution: Moore's law[2] suggests that computational power has increased hundredfold since the paper's publication in 2000. Additionally, many devices are largely linear at low volumes, so the majority of impulses could be taken at high amplitudes. If different sets of measurements are taken for various device settings, the interpolation extends to a problem in multidimensional parameter space.

While "Dynamic Convolution" presents a novel way of treating non-linear systems, it's its inability in recreating time-varying behaviour that ultimately makes the method unsuitable for high-quality vacuum tube simulation. However, others have admitted [7] that the concept of sampling at multiple amplitudes may increase the quality of some of the non-linear with memory methods.

2.1.3 Non-Linear Systems with Memory

The additional requirement of memory adds another level of complexity to the system. This is the entry point for the original work in this project. The majority of techniques for this class of system exploit the Volterra series in some form.

[1] As positive and negative values are included, we end up with the range $-2^{16}/2$ to $2^{16}/2 - 1$. Digital signals generally also include some headroom, so F_{sv} is generally even less than half of 2^{16}.

[2] Moore's Law states that computational power approximately doubles every 18 months.

Volterra Series In its most general form, the Volterra series is given as [8]:

$$y(t) = k_0(t) + y_1(t) + y_2(t) + \cdots \tag{3}$$

$$y_n(t) = \int \cdots \int k_n(\tau_1, \tau_2, \cdots, \tau_n) x(t-\tau_1) x(t-\tau_2) \cdots x(t-\tau_n) d\tau_1 d\tau_2 \cdots d\tau_n \text{ for } n > 1$$

The term k_n is referred to a the n^{th} order Volterra kernel. As we shall see later, these are closely related to the higher order impulse responses, but are not exactly equivalent.

In this form however, Eq. (3) is prohibitively complex to be of any practical use. Farina claims [9] that at least a 5th order representation is required for an accurate representation. This is simply not possible using the full form of Eq. (3) with today's computing power, as the n^{th} Volterra kernel is n dimensional!

However, by treating the system as a Hammerstein model, where we assume the signal path contains a memoryless non-linearity followed by a linear part with memory, we can simplify the expression considerably by losing cross terms (i.e. for $n > 1$, $\tau_n = \tau \; \forall n$) [9]. This process is often referred to as taking the Diagonal Volterra coefficients:

$$y(t) = k_0 + k_1(t) \otimes x(t) + k_2(t) \otimes x(t)^2 + \cdots + k_M(t) \otimes x(t)^M \tag{4}$$

Mathematically this can be interpreted as assuming all non-diagonal terms are zero. This is generally not the case, although often off-diagonal values are small enough not to have too large an effect. The fact that the problem now only scales linearly with M, generally outweighs this information loss.

In the memoryless case, the representation simplifies even further [10]:

$$y(t) = a_0 + a_1 x(t) + a_2 x(t)^2 + \cdots + a_M x(t)^M \tag{5}$$

From this, only the set of constant coefficients $\{a\}$ are needed to represent the system.

Using Pseudo-Random Noise In his 2005 paper [10], Hawksford outlines a method for determining the set of kernels in Eq. (4), $k_n(t)$. Confusingly, Hawksford uses the symbol h_n to describe these kernels - these should not be confused with the n^{th} order impulse response (these *are* related, see next section on Sine Sweeps). For ease of comparison with the original paper, we will temporarily adopt this notation.

As the set of kernels $h_r(n)$ are independent, we also require M independent input signals to determine them: a set of simultaneous equations needs M knowns to find M unknowns. For this, Hawksford suggests using a set of pseudo-random noise sources. However care must be taken when generating these inputs: precision issues with small number division can be avoided by generating the random number sequence with a complex exponential. Given an initial, uniformly distributed sequence of random numbers, $p(n)$ with Fourier Transform $P(n)$, we find input vector X by:

$$X = e^{i \arg(P(n))}$$

$$\text{or } x = \text{Re}(\text{IFT}\{X\}) \text{ in the time domain}$$

Hawksford uses the well established technique of Fourier Transforming into the frequency domain for the heavy calculations; the set of computationally expensive convolutions in (4) now become multiplications. To fully determine the set of M kernels, we must solve the set of M simultaneous equations (for independent inputs $r = 1 \ldots M$).

$$Y_r = H_0 + H_1 X_{r,1} + H_2 X_{r,2} + \cdots + H_M X_{r,M}$$

where $H_r = \text{FT}\{h_r(n)\}$, $Y_r = \text{FT}\{y_r(n)\}$, and $X_{r,q} = \text{FT}\{x_r(n)^q\}$. Subtracting off the constant H_0 term, we can instead rewrite this in following matrix form:

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

Inversion of matrix $\mathbf{X}$ allows the set of kernels to be found as:

$$H = \mathbf{X}^{-1} Y$$
$$h = \text{IFT}\{H\}$$

The inversion will be computationally expensive (the matrix is 3-dimensional and can typically have of the order of 2^{20} elements!), but as this is only dependent on the noise input vectors X, it can be found at compile time and may be reused for multiple system measurements. Tests by the author on known, virtual systems proved successful: for a sample memory-less system, the set of coefficients $\{a\}$ (Eq. (5)) were recovered. A system with memory (set of lowpass filters) was similarly tested with positive results. Further testing on real systems is required though for proper verification of the method.

2.1.4 Exponential Swept Sine Signal

In their 2010 paper [11], Novak et. al suggest an alternative method using sine wave signals of varying frequency, based on an original technique by Farina et. al [12]. By using a swept sine signal with an exponentially increasing instantaneous frequency[3], it is possible to separate out the various impulse responses corresponding to different orders of harmonic distortion, i.e. at different integer multiples of the fundamental frequency. For a frequency range $[f_1, f_2]$, the signal can be defined as:

$$x_s(t) = A_s \sin\left\{2\pi L \left[\exp\left(\frac{f_1 t}{L}\right) - 1\right]\right\}$$

where for a input signal of time length $\hat{T}$:

[3] Alternatively known as a chirp

$$L = \text{Round}\left(\frac{\hat{T} f_1}{\ln(f_1/f_2)}\right)$$

We can define an "inverse filter" $\tilde{x}_s(t)$ of signal $x_s(t)$, i.e. a time reversed version with an additional amplitude modulation so that $x_s(t) \otimes \tilde{x}_s(t)$ is the Dirac delta function, $\delta(t)$:

$$\tilde{x}_s(t) \propto \exp(-t) x_s(-t)$$

Thus, convolution with the output produces a superposition of the set of impulse responses $\{h(t)\}$, or equivalently a set of frequency responses $\{H\} = \text{FT}\{h(t)\}$.

$$y(t) \otimes \tilde{s}(t) = \sum_{n=1}^{\infty} h_n(t + \Delta t_n) \tag{6}$$

As these responses are separated by a determinable time Δt_n (see Fig. 2), these can individually identified.

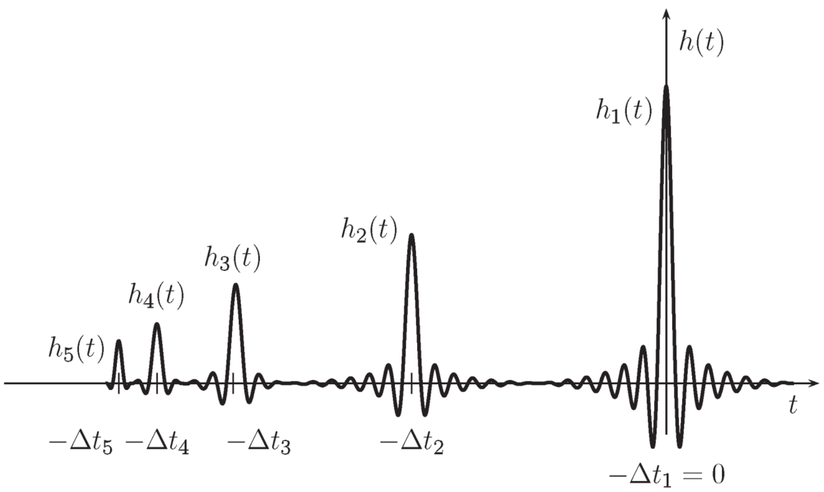
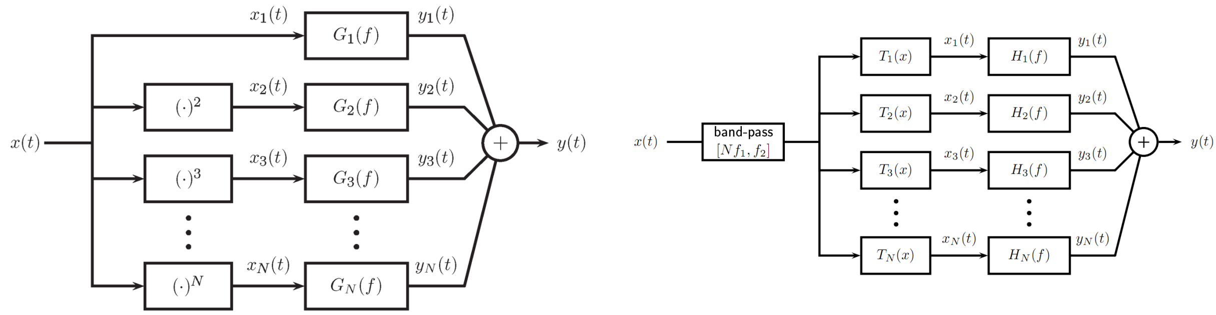


Figure 2: Separation of impulses of increasing harmonic order. Figure from [11].

Finally, we can calculate the diagonal Volterra kernels. In the frequency domain it can be shown analytically that these kernels $G_n(f)$ can be formed by a linear superposition of frequency responses $H_n = \text{FT}(h_n)$. There isn't a direct mapping of H_n to G_n as $\sin(x)^n$ is not exactly the same as the n^{th} harmonic, $\sin(nx)$.

The details of this process are beyond the scope of this review however, for further reading see [11]. As in Eq. (4), these kernels (which can be also though of as filters) are applied to copies of the input signal $x(t)$, raised to increasing powers, in a model named the Generalized Polynomial Hammerstein (GPH) model. This is perhaps best understood by reference to Fig. 3a.



(a) Schematic of GPH Model (figure from [11]) (b) Chebyshev Non-Linear Model (taken from [7])

Figure 3: Various models for treating non-linearities

Chebyshev Non-Linear Model In a follow-up paper [7], Novak et. al present an alternative method of using the set of frequency responses $\{H\}$ obtained in Eq. (6). The responses are left as is, instead opting to pass each branch of the input through a *Chebyshev Polynomial of the First Kind*, $T_n(x)$, schematically depicted in Fig. 3b. The series is defined as:

$$
\begin{aligned}
T_1(x) &= 1, \\
T_2(x) &= x, \\
&\vdots \\
T_n(x) &= 2xT_{n-1}(x) - T_{n-2}(x)
\end{aligned}
$$

The full details of the method can be found in [7], but essentially the technique uses the property that T_n produces additional higher harmonics of order n (for a cosine input), as is seen in harmonic distortion. Many similarities exist between this and the GPH model; the main difference is the choice of mathematical series used.

2.2 White Box Approaches

Often when we have knowledge of the internal workings of a system, it can be worthwhile exploring so-called White Box approaches. These methods build a model of the system based on previous knowledge. First we review waveshaping techniques, then briefly look at attempts at modelling valve amplifier circuitry.

2.2.1 Waveshaping

If the system is non-linear but memoryless, one of the simplest methods for recreating distortion effects is using a static waveshaper. This is simply just a map of input levels to output levels. This can be achieved by applying a function describing the non-linearity, for example Suyama and Araya describe a cubic function in their patent [13]:

$$y = \frac{3x}{2}\left(1 - \frac{x^2}{3}\right)$$

As the function is largely linear for most operating amplitudes, the authors suggest passing the signal through several times. An alternative patent by Doidic [14] uses a piecewise asymmetric function to produce both odd and even harmonics which is more typical of valve behaviour [2]. These methods are commercial by design however, and aim more to roughly emulate guitar amp sound than scientifically determine it; they generally sacrifice accuracy for efficiency. Further speed gains are proposed whereby input-output mappings are stored beforehand in lookup tables, rather than calculated on the fly, for example [15].

2.2.2 Circuit Modelling

An altogether different approach is to model the circuit in a component by component basis. Attempts have been made [16, 17] using open source circuit prototyping software SPICE[4] [18], with somewhat encouraging results, although so far the work has largely been non-realtime. The authors also stress that the quality of simulation is only as good as the quality of individual components in the software.

Recent attempts by Pakarinen [19] propose using Digital Wave Filters (DWFs) to represent the circuits. In general, circuit components can act bi-directionally which adds a level of complexity. The DWF process determines the resulting ODE/PDEs that describe the system - these generally must be solved numerically. For an introduction to DWFs, see [20].

3 Outlook

So where does this project fit in? Clearly there is still much progress to be made in the field. The latest thinking suggests that in fact adaptive models may be the way forward. This methodology sets a trial model for the system and then quantitatively compares simulated and actual outputs. The model is then refined, and the process repeated until an acceptable tolerance is reached.

Existing proposals of this class include a type of genetic algorithm [21], whereby model selection is based on “evolutionary fitness” - models that are closer to the desired output have greater weight, whereas weak fits are given a genetic weakness. The model (unsurprisingly) works in much the same way as genetic evolution in the natural world [22, 23], with crossover, mutations and a certain amount of randomness.

There is also much room for exploration of White Box methods. While these methods are difficult to apply to non-linear systems in general, guitar valve amplifiers have a specific circuitry and design more suited to this approach. This project hopes to investigate the effectiveness of component by component modelling, although the results may be somewhat limited by the quality of existing physical models for component parts.

[4] Simulation Program with Integrated Circuit Emphasis

Finally, the project aims to look at the important field of optimisation. Many of the models reviewed here are mathematically complete solutions to the problem; the issue being the theoretically "infinite" run times required to find them. A large part of computational physics is about making sensible approximations in these cases; this project hopes to evaluate the optimal point between performance and accuracy. By analysing some of the heavier matrix operations involved, it is hoped that new optimisations may be presented to allow more efficient computation.

4 Conclusions

Digital simulations of valve amplifiers will likely never become *fully* accepted by the music production community. DSP technology has, often unfairly, a certain stigma attached to it, suffering to what is sometimes dubbed "the fetishisation of analogue", where people may flat out ignore digital possibilities, opting for a purely analogue philosophy.

However this review should demonstrate that the field is a very active area of research. Existing models have analytically been shown to be very close to real system behaviour, and many have had favourable results from initial subjective tests by guitarists.

This project hopes to built on the existing foundation of work to push the accuracy with which non-linear systems with memory can be represented, within the practical constraints of current computing technology. As techniques and hardware capabilities improve, as the model tends towards the reality, perhaps we may seen even the most dedicated analogue audiophiles accept digital alternatives.

References

[1] Google shopper: Valve amplifiers. `http://www.google.co.uk/products?q=valve+guitar+amplifier&scoring=pd`, Retrieved 2010.

[2] Jyri Pakarinen and David T. Yeh. A review of digital techniques for modeling vacuum-tube guitar amplifiers. *Comput. Music J.*, 33(2):85–100, 2009.

[3] H. Kuttruff. *Room acoustics*. E-Libro. Spon Press, 2000.

[4] B. Venkataramani and M. Bhaskar. *Digital signal processors: architecture, programming and applications*. Tata McGraw-Hill, 2002.

[5] Michael J. Kemp. Analysis and simulation of non-linear audio processes using finite impulse responses derived at multiple impulse amplitudes. In *Audio Engineering Society Convention 106*, 5 1999.

[6] Liquid mix: Dynamic convolution. `http://www.focusrite.com/products/liquid/liquid_mix/`, Retrieved 2010.

[7] A. Novak, L. Simon, J. Gilbert, and P. Lotton. Chebyshev model and synchronized swept sine method in nonlinear audio effect modeling. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, 2010.

[8] Martin. Schetzen. *The Volterra and Wiener theories of nonlinear systems / Martin Schetzen.* Wiley, New York :, 1980.

[9] Adriano Farina, Angelo; Farina. Real-time auralization employing a not-linear, not-time-invariant convolver. In *Audio Engineering Society Convention 123*, 10 2007.

[10] Malcolm J. Hawksford. System measurement and identification using pseudorandom filtered noise and music sequences. *J. Audio Eng. Soc*, 53(4):275–296, 2005.

[11] A. Novak, L. Simon, F. Kadlec, and P. Lotton. Nonlinear system identification using exponential swept-sine signal. *Instrumentation and Measurement, IEEE Transactions on*, 59(8):2220 –2229, aug. 2010.

[12] Angelo Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Engineering Society Convention 108*, 2 2000.

[13] A. Suyama and T. Araya. Sound effector capable of imparting plural sound effects like distortion and other effects. *US Patent*, 1996.

[14] M. et al Doidic. Tube modelling programmable digital guitar amplification system. *US Patent No. 5789689*, Issued 1998.

[15] G Kramer. Digital signal processor for providing timbral change in arbitrary audio and dynamically controlled stored digital audio signals. *US Patent No. 4991218*, Issued 1991.

[16] W. M. Leach. Spice Models For Vacuum-Tube Amplifiers. *JOURNAL OF THE AUDIO ENGINEERING SOCIETY*, 43(3):117–126, MAR 1995.

[17] Improved vacuum tube models for spice simulations. `http://www.normankoren.com/Audio/Tubemodspice_article.html`, Retrieved 2010.

[18] Spice: General-purpose circuit simulation. `http://bwrc.eecs.berkeley.edu/Classes/IcBook/SPICE/`, Retrieved 2010.

[19] J. Pakarinen and M. Karjalainen. Enhanced Wave Digital Triode Model for Real-Time Tube Amplifier Emulation. *IEEE TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING*, 18(4):738–746, MAY 2010.

[20] Physical audio signal processing: Wave digital filters. `https://ccrma.stanford.edu/~jos/pasp/Wave_Digital_Filters.html`, Retrieved 2010.

[21] Yen-Wei Chen, S. Narieda, and K. Yamashita. Blind nonlinear channel identification based on constrained hybrid genetic algorithm. In *Instrumentation and Measurement Technology Conference, 2001. IMTC 2001. Proceedings of the 18th IEEE*, volume 2, pages 1253 –1257 vol.2, 2001.

[22] M. Mitchell. *An introduction to genetic algorithms.* Complex adaptive systems. MIT Press, 1998.

[23] J.-M. Renders and S.P. Flasse. Hybrid methods using genetic algorithms for global optimization. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 26(2):243 –258, apr. 1996.