NONLINEAR FILTERS

Risto Holopainen
NoTAM
Norwegian Network for
Technology, Acoustics and
Music

ABSTRACT

Many distortion effects can be implemented as nonlinear (NL) filters. We review some common NL-filters, and introduce new ones that combine features from linear filters and waveshaping. The great interest in recent years in analog effects has resulted in several attempts at digital emulations of classic analog products. However, many interresting digital effects can be construed without knowledge of analog counterparts. Although NL-filters seem to require complicated mathematical analysis, much insight can be gained simply by carrying out some relevant experiments.

1. INTRODUCING NONLINEAR FILTERS

According to a negative definition, a nonlinear filter is any filter that does not meet the criteria of linearity. For a linear system $\mathfrak{L}$, given the two signals $\{x\}$ and $\{y\}$ and a constant a, two familiar conditions must hold:

$$\mathfrak{L}(ax) = a\mathfrak{L}(x) \quad (1)$$

$$\mathfrak{L}(x) + \mathfrak{L}(y) = \mathfrak{L}(x + y) \quad (2)$$

In contrast, we will describe several filters that share the property that these two conditions are not met. They can be written in the general form

$$y_n = f(x_n, \ldots x_{n-N}, y_{n-1}, \ldots, y_{n-M}) \quad (3)$$

where $f(\cdot)$ is a nonlinear function. Dynamic range processing is a common type of nonlinear processing.

An important class of nonlinear filters are the statistical filters, e.g. the median filter, which require sorting of its input. However, most of the NL-filters we will consider here combine a regular linear filter with distortion or waveshaping The waveshaper W and the linear filter F can be cascaded in two obvious ways, or nested in a recursive structure, or they can be combined in parallel structures.

Both of the two cascaded structures can be understood by analyzing the behaviour of their constituent parts, if they are known. A typical example could be a lowpass filter followed by distortion. There is always a risk that the nonlinearity introduces an objectionable amount of aliasing. If the distortion function is a polynomial of degree P, and the lowpass effectively cuts frequencies above 1/P of the Nyquist frequency, then this effect is guaranteed to produce band limited results and aliasing can be avoided [12]. In cases where the waveshaping function does not produce band limited output, oversampling may be a solution.

Much more intriguing, and often sonically rewarding results can be obtained with a feedback structure. In particular, some filters may exhibit chaos. A necessary, but not sufficient requirement for chaos to occur, is that the system is recursive, and has a nonlinearity in the feedback path.

In the next section we discuss some standard modelling approaches to nonlinear systems, and give examples of previous work on NL-filters. Then we propose a few new variations on NL-filters. In particular, some modifications of a bandpass filter are studied. New digital effects can be constructed in a process similar to analysis by synthesis.

2. MODELLING APPROACHES

Nonlinear time invariant systems with memory can be modelled by means of convolution. Two common approaches are the Volterra series representation and dynamic convolution.

2.1. Volterra Series

Some nonlinear systems can be described in terms of their Volterra series expansion. A first order Volterra system is simply a one-dimensional convolution, $y_n = h_n * x_n$. A second order Volterra system also involves second degree cross terms,

$$y(n) = h_1(n) * x(n) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_2(i, j) x(n-i) x(n-j) \quad (4)$$

and the third order system has third degree terms, and so forth. The system is described in terms of its present and past input, and does not depend on its previous output. Thus a nonlinear recursive filter is not ideally suited for this model; it would take an infinite order Volterra system to represent it. Furthermore, if the nonlinearity involved is not a low degree polynomial, the complexity increases even more. For instance, the very useful tanh function would have to be represented as an infinite Taylor series expansion.

However, an infinite degree polynomial distortion function would yield aliasing, which can be avoided by truncating its Taylor series. Such an approach has been taken in modelling a Moog filter [6].

There are methods to analyze analog devices and identify their Volterra kernels. While these tend to be of a very high order, they can be reduced to a form that is more suitable for digital implementation [10].

A simplified form of the Volterra series is obtained by discarding all multiplications with delayed samples; i.e., the system is modelled as a sum of linear convolutions of x, x^2, ..., x^p. This corresponds to processing with nonlinear distortion followed by a linear filter, and has been used for loudspeaker simulation [4].

2.2. Dynamic Convolution

Dynamic convolution was introduced as a means to simulate analog effects, with the additional benefits of reduced noise levels and flexibility of operation that comes along with its digital implementation. This method was pioneered and patented by Michael Kemp [8], but an almost identical approach was being pursued at roughly the same time by Angelo Farina, who called it impulse response switching [3]. Although the purpose was to simulate analog processing, the algorithm could be used with any set of impulse responses, which do not need to be sampled from an analog system.

In systems consisting of a nonlinearity followed by a linear filter, the impulse response will depend on the amplitude. The simulation process begins with sampling a collection of impulse responses at several different amplitude levels and normalizing their peak amplitudes. Let $h_k[n]$ be the impulse response in the kth amplitude interval. Then, for each input sample $x[n]$ an impulse response will be chosen. Let $S(x[n])$ be the function that assigns an impulse response corresponding to a given amplitude. The output $y[n]$ is a summation of convolutions with kernels, that, in general, will vary from one sample to the next:

$$y[n]=\sum_{m=0}^{M-1} x[n-m]h_{S(x[n-m])}[m] \qquad (5)$$

If the number of impulse responses is lower than the number of discrete amplitude steps (e.g. 256 in an 8-bit signal), they should be interpolated to match variations in amplitude. In Kemp's version, the amplitude steps are regularly spaced on a linear scale, whereas Farina used logarithmic spacing [3].

The quality of dynamic convolution depends on the number of impulse responses and their length. While the purpose is to introduce some amount of distortion, having too few impulse responses will most likely introduce excessive distortion. As a general guideline for designing new dynamic convolution effects from scratch, the impulse responses should change very little from one amplitude level to the next.

3. EXAMPLES OF NL-FILTERS

A wide assortment of NL-filters has been developed to meet various needs in image and audio processing. One common type is the median filter. Its implementation and use differ from other NL-filters in many respects. Another broad class comprises the waveguide filters, which are basically a recurrent structure, including a delay. There are also lots of attempts at modelling classic analog filters, such as the Moog filter [6, 7].

3.1. Ordinary median filters

Most often encountered in image processing, the two-dimensional median filter is defined as the median of a set of pixel values. The one-dimensional case can be used for audio filtering. An Nth order median filter sorts the last N input samples and returns the middle value.

The median filters' purpose is to remove impulsive noise, i.e., extreme peak values. An improved (but far from perfect) peak remover can be built of a differentiator, followed by a median filter, and a leaky integrator to compensate for differentiation. While a median filter distorts a sine wave, it also attenuates high frequency content of white noise, in effect acting as a lowpass filter.

Median filters have some unusual properties: its impulse response is always zero; and every output value is guaranteed to be identical to one of the input values. This may be useful in applications such as algorithmic composition. Consider a signal representing a discrete set of pitches. A median filter will never output any intermediate values; hence rounding, and pitches outside the scale system, is avoided.

3.2. Recursive and weighted median filters

Several variations of the standard median filter have been suggested. The median filter can be made recursive by feeding back one or more of its last output values. A recursive median filter

$$y_n = median(x_n, x_{n-1}, ..., x_{n-N}, y_{n-1}, ..., y_{n-M}) \qquad (6)$$

may be more effective than the non recursive filter of same order. This can lead to a great efficiency improvement, since most sorting algorithms have an $O(n)$ complexity at best. The number of feedback taps y_n should not exceed the number of direct inputs x_n; otherwise the median filter will stick to one of its values and keep repeating it forever.

Even more exotic are the median filters with weighed coefficients, allowing negative weights [11]. A sample value is repeated a number of times, making its selection more probable the more it is repeated. For negative weights, the sign of the sample value is inverted. Such filters may be designed to have a bandpass or highpass characteristic on white noise, combining the frequency selectivity of linear filters with the ability to filter out impulsive noise.

3.3. The Dobson-ffitch nlfilt

The Csound opcode nlfilt is an implementation of the filter structure

$$y_n = x_n + a y_{n-1} + b y_{n-2} + d\, y_{n-L}^2 - C \quad (7)$$

as originally proposed by Dobson & ffitch [2]. For large parameter ranges, it sounds much like a comb filter, and in regions of self-oscillation it may be reminiscent of acoustic feedback. Most importantly, it easily becomes unstable.

It's tempting to amend the filter and counteract its instability. One could wrap the feedback path in a clipping function. Clearly this makes the occurrence of instability impossible for any parameter ranges. Another trick is to put a compressor inside the filter, and reduce the gain whenever y_n exceeds some threshold. While this is a worthwhile strategy in other cases [1], here it is not. Unfortunately, the Dobson-ffitch filters' sonic potential is reduced by any tampering with its structure. Its most interresting sound is often to be found in regions tantalizingly near instability.

3.4. The Babylonian square root filter

One of the oldest known algorithms, is the Babylonian square root algorithm. Despite its age, its application as an audio filter seems to be novel. Let $x_n > 0$ be a constant signal (i.e. $x_n = a$ for all n) and $y_0 = 1$. Then

$$y_n = \frac{1}{2}\left(y_{n-1} + \frac{x_n}{y_{n-1}}\right) \quad (8)$$

will rapidly converge to the square root of a.

In practice, the signal x_n is first mapped into the positive interval (0, P]. Then the filter is applied, and finally the signal is mapped back into the interval [-1,1]. The sound of this filter is a mild distortion. Its asymmetric character introduces harmonic distortion on both odd and even partials. The impulse responses are generally characterized by very short transients.

A comparison with the square root function used as a waveshaper reveals that the Babylonian algorithm produces almost the same waveshape. In both cases, a sine wave becomes pointed in its negative half, while the positive part is more rounded. Obviously, the difference stems from the inertia, the time it takes for the algorithm to converge to its limit. Since this filter does not produce band limited output, some oversampling may be necessary.

4. DISTORTED BANDPASS FILTERS

Taking a bandpass filter as a point of departure, we experiment with nonlinear alterations. Their effects range from distortion, to ringing and chaos.

4.1. Variations on filter structure

Among several possible designs of a bandpass filter, we choose

$$y_n = G(x_n - R x_{n-2}) + b_1 y_{n-1} + b_2 y_{n-2} \quad (9)$$

where the parameters G, R, b_1 and b_2 are calculated from the center frequency f_c, bandwidth and sampling frequency [9].

This filter structure can be modified by inserting a nonlinear, bounded function f, for example:

$$y_n = f\{G(g x_n - R x_{n-2}) + b_1 y_{n-1} + b_2 y_{n-2}\} \quad (10)$$

An additional parameter g controls the input gain. This formulation is guaranteed to be stable and to produce an output with the same bound as f.

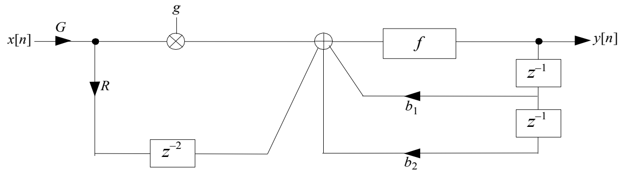

Figure 1. Second order filter structure with one nonlinearity.

Another filter structure that has been tried out is the one in figure 2. This filter structure produces markedly different sounds from the one above. For $g > 1$

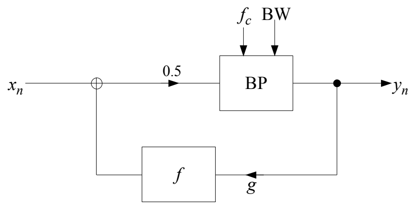

Figure 2. Recursive bandpass and limiting function.

it will tend to ring on the center frequency. However, very little distortion is introduced. For all these filter structures, the choice of limiting function may have a significant impact on the filter's behaviour.

4.2. Choice of limiting function

The tanh function seems to prevail as limiting function. Indeed, it has many desirable properties: Its range is bounded to the interval (-1,1); it is continuous and monotonously increasing, and almost linear near 0.

Other functions we have tried include arc tan, soft clipping [12], hard clipping, sine, and the lowered bell function:

$$f(x) = \frac{2}{x^2 + 1} - 1 \quad (11)$$

When these nonlinearities appear inside filters, it is in many respects advantageous that they approximate the identity function $f(x) = x$ in the lower amplitude range, as this will ensure that the filter behaves more or less like a linear filter when the signal's amplitude is low.

An important implication is, that the frequency response as calculated for the corresponding linear filter is valid as an estimation when the amplitude and gain settings are low. Clearly the lowered bell function violates these demands, in particular as $f(0) = 1$.

4.3. Chaotic filters

In accordance with a standard definition of chaos, a filter would be recognized as chaotic if two of its impulse responses diverge exponentially over time when the impulse amplitude is slightly changed. This is of course what is known as sensitive dependence on initial conditions, and the divergence rate is measured by the lyapunov exponent [5]. Chaotic filters can be analyzed by the standard methods of chaos theory, such as the bifurcation diagram and lyapunov exponents. But since these analysis methods actually only deal with impulse responses, they cannot tell the whole story of the filter.

A one-dimensional map can be chaotic only if it is not invertible. Consequently, putting a regular monotone limiting function into the filter may not make it chaotic. The Dobson-ffitch filter has chaotic regions precisely because of the square function.

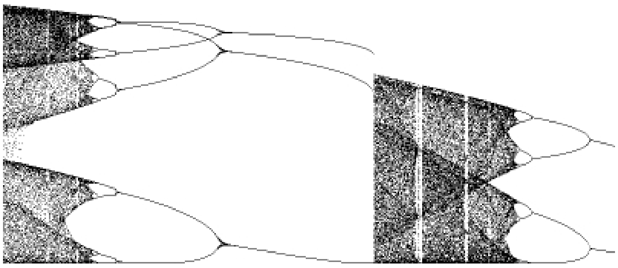

Figure 3. Bifurcation diagram of eq. 10 with bell function. Gain and Q-factor are both 1, frequency goes from 0 to fs/8 on the right.

The filter structure of figure 1 (eq. 10) in combination with eq. 11 has a period doubling rout to chaos, as shown in fig. 3. Here, the varying parameter is center frequency. Going from high gain settings and decreasing the gain, as well as increasing the Q-factor, may also result in period doubling bifurcations.

Another combination of filter structure and limiting function that yields remarkable results is the recursive structure (fig 2) and sine function. When processing an input sound, it produces a mellow noisy tone for a wide range of parameter settings.

When evaluating these chaotic filters, it is a striking fact that the orbits have the typical characteristic of iterated maps, that of bifurcations or period doubling. This means, that the filter will ring on frequencies that divide the sampling frequency, typically quite high frequencies.

5. CONCLUSIONS

Nonlinear filters is a vast topic, with its two main areas being statistical filters and distortion with memory. Volterra series and dynamic convolution are two frequently used models for simulating analog systems.

Although linear filter theory is not directly applicable to nonlinear filters, some of its methods may be used if viewed with proper suspicion. It may be revealing to investigate impulse responses at various amplitudes, as well as responses to other signals with a flat spectrum. Any problems with aliasing are easily heard when filtering a sine tone, slowly swept from Nyquist to DC. However, the filter's value as a musical effect will only become apparent when tested with musical signals. Often the sound of a recursive nonlinear filter is more fascinating than those that can be produced by structures separated into a waveshaping and linear filter part.

REFERENCES

[1] Berdahl, E. & Smith, J. O. (2006). Some Physical Audio Effects. Proc. of the 9th Int. Conference on Digital Audio Effects. Montreal, Canada.

[2] Dobson, R. & ffitch, J. (1996). Experiments with Non-Linear Filters. Discovering Excitable Regions. ICMC Proceedings. Hong Kong, China.

[3] Farina, A. & Armelloni, E. (2005). Emulation of Not-Linear, Time-Variant Devices by the Convolution Technique. AES Italian Section, Annual Meeting. Como, Italy.

[4] Farina, A, Bellini, A, Armelloni, E. (2001). Non-Linear Convolution. A New Approach for the Auralization of Distorting Systems. Proceedings of the 110th AES Convention. Amsterdam, the Netherlands.

[5] Frøyland, J. (1992). Introduction to Chaos and Coherence. Bristol: Institute of Physics Publishing.

[6] Hélie, T. (2006). On the Use of Volterra Series for Real-Time Simulations of Weakly Nonlinear Analog Audio Devices: Application to the Moog Ladder Filter. Proc. of the 9th Int. Conference on Digital Audio Effects. Montreal, Canada.

[7] Huovilainen, A. & Välimäki, V. 2005. New Approaches to Digital Subtractive Synthesis. Proceedings of the ICMC 2005. Barcelona, Spain.

[8] Kemp, M. 1999. Analysis and Simulation of Non-Linear Audio Processes using Finite Impulse Responses Derived at Multiple Impulse Amplitudes. The 106th AES convention. Munich, Germany. Preprint number 4919.

[9] Moore, F. R. (1990). Elements of Computer Music. New York: Prentice Hall.

[10] Schattschneider, J. & Zölzer, U. (1999). Discrete-Time Models for Nonlinear Audio Systems. Proc. of the 2nd Workshop on Digital Audio Effects. Trondheim, Norway.

[11] Shmulewich, I. & Arce, G. (2001). Spectral Design of Weighted Median Filters Admitting Negative Weights. IEEE Signal Processing Letters, vol 8, no 12.

[12] Zölzer, U. 2002. DAFX.– Digital Audio Effects. John Wiley & Sons.