ESTIMATING THE AMPLITUDE OF THE CUBIC DIFFERENCE TONE USING A THIRD-ORDER ADAPTIVE VOLTERRA FILTER

Kyogu Lee, Minjong Kim

Center for Computer Research in Music and Acoustics
Music Department, Stanford University
{kglee,minjong}@ccrma.stanford.edu

ABSTRACT

Design method of a nonlinear filter to estimate the amplitudes of cubic difference tones is presented. To this end, a third-order Volterra filter is used to model the nonlinearity of our auditory system, and the filter coefficients are obtained using an adaptive process. The results show the filtered outputs follow very closely the experimental data as the intensity levels and the frequencies of inputs vary especially when the frequency separation between the two primary tones is not large.

1. INTRODUCTION

When there is an input stimulus with two sinusoidal components called *primary tones* at frequencies f_1 and f_2 ($f_1 < f_2$), our ear not only perceives tones at frequencies f_1 and f_2, but also hears many other distortion tones at frequencies $k_1 f_1 \pm k_2 f_2$, $k_i \in Z$, especially when the intensity level of the input is high. These distortion tones, called *combination tones*, are due to nonlinear behavior in our auditory system. Such nonlinearity in the ear was reported as early as 1856 by von Helmholtz, when he wrote a theory on combination tones using a power series model [1].

Among many possible combination tones, the most audible and prominent ones are cubic difference tones or the CDTs at frequency $2f_1 - f_2$, and quadratic difference tones (QDTs) at frequency $f_2 - f_1$. While many experiments show that the quadratic difference tone level observes the classical square-law in general, the amplitude behavior of the CDTs is so unusual that every experiment yields different results. Such unusual behavior of the CDTs has caused a great deal of research on modeling auditory nonlinearity with an attempt to match experimental data as closely as possible. Zwicker, who was the first to observe the abnormal behavior of the amplitude of the CDTs, performed extensive experiments using a cancellation tone and proposed a model with a nonlinear feedback loop [2]. Goldstein presented methods for measuring more precisely the intensity of the CDTs and difference tones by presenting a probe tone to facilitate adjustment of the cancellation tone, and suggested the idea of *essential nonlinearity*, according to which the relative level of the CDT is almost independent of the stimulus level [3]. Smoorenburg observed that the CDTs can be heard at low stimulus levels only in a restricted region below f_1, the so called "audibility region of combination tones" [4].

In the present paper, the authors present design methods for a nonlinear filter that better fits various experimental data for the cubic difference tones. To accomplish this, a third-order Volterra filter is used to model the nonlinearity in our auditory system, and the filter coefficients are estimated based on the adaptive process with a third sinusoidal primary tone added to an input. In order to validate the model, arbitrary input stimuli of different frequency interval and/or different primary tone levels are fed to the filter, and the results are compared with the experimental data.

In the next section, the basic framework of a Volterra filter for nonlinear system identification is reviewed. In the following section, the authors present an algorithm to obtain the filter coefficients using an LMS-based adaptive process by carefully choosing a desired response of the system, followed by the simulation results and comparison with the experimental data.

2. VOLTERRA FILTERS FOR NONLINEAR SYSTEMS

The nonlinear behavior of the hearing mechanism proposed by von Helmholtz [1] was later simplified by Fletcher [5] using a classical power series represented by

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n, \tag{1}$$

where x and y are the input and output of the system, and a_n are constants. The nonlinearity in the power series represented by Equation (1) is generally considered to increase as input level increases.

A nonlinear system with memory represented by means of an extension of power series, known as *Volterra series expansion*, can be described as

$$y(t) = h_0 + \sum_{p=1}^{\infty} \overline{h}_p[x(t)], \tag{2}$$

where $x(t)$ and $y(t)$ are the input and output signals, and

$$\begin{aligned} \overline{h}_p[x(t)] \quad &= \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_p(\tau_1, \cdots, \tau_p) x(t - \tau_1) \\ \cdots \quad & x(t - \tau_p) d\tau_1 \cdots d\tau_p. \end{aligned} \tag{3}$$

The multidimensional functions $h_p(\tau_1, \cdots, \tau_p)$, called the *Volterra kernels*, can completely characterize nonlinear systems representable by Volterra series such as polynomial systems [6].

In a manner similar to the continuous case, an N-th order discrete-time Volterra filter for the causal, nonlinear system with memory length M can be described by

$$y(n) = h_0 + \sum_{p=1}^{N} \overline{h}_p[x(n)], \tag{4}$$

where

$$\overline{h}_p[x(n)] = \sum_{m_1=0}^{M} \cdots \sum_{m_p=0}^{M} h_p(m_1, \cdots, m_p) \times x(n-m_1)\cdots x(n-m_p). \quad (5)$$

Since we are interested in modeling the nonlinear behavior of cubic difference tones, we use a truncated Volterra filter with the order $N = 3$, which is defined as:

$$\begin{aligned} y(n) &= \sum_{m_1=0}^{M-1} h_1(m_1)x(n-m_1) \\ &+ \sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_2(m_1,m_2)x(n-m_1)x(n-m_2) \\ &+ \sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1}\sum_{m_3=0}^{M-1} h_3(m_1,m_2,m_3)x(n-m_1)x(n-m_2)x(n-m_3), \end{aligned} \quad (6)$$

where M is the memory length of the filter.

In the next section, we describe how to estimate filter coefficients by employing the least-mean-squares (LMS) algorithm.

3. ESTIMATING FILTER COEFFICIENTS USING THE LMS ALGORITHM

3.1. The LMS Algorithm for Third-Order Volterra Filter

There are several ways to estimate the Volterra filter coefficients, among which the minimum mean-square error (MMSE) estimation and the least-squares estimation are well known algorithms of direct estimation schemes. In situations where the nonlinear system to be modeled is time-varying and/or the statistics of the signals involved are not known a priori, adaptive filters could be useful while direct estimation methods may fail [6]. We chose an adaptive algorithm so as to make the model more viable in such situations.

Among many adaptive algorithms, the least-mean-square algorithm, or LMS algorithm makes use of a special estimate of the gradient, and is represented by the following two equations [7]:

$$\epsilon(n) = d(n) - \mathbf{X}^{\mathrm{T}}(n)\mathbf{W}(n), \quad (7)$$

and

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) - \mu\hat{\nabla}(n) \\ &= \mathbf{W}(n) + 2\mu\epsilon(n)\mathbf{X}(n), \end{aligned} \quad (8)$$

where $\epsilon(n)$ is the error signal, $d(n)$ is the desired response, $\mathbf{X}(n)$ is the input vector, $\mathbf{W}(n)$ is the weight vector at time index n, and μ is the gain constant that regulates the speed and the stability of adaptation. Figure 1 shows such an adaptive linear combiner with the memory length M.

The LMS algorithm for linear adaptive filtering can be easily expanded to implement the third-order adaptive Volterra filter since the filter output defined in Equation (6) is simply the linear combination of its linear input signals and the second and third order cross-products of its linear input signals.

The third-order Volterra filter can be represented in vector form as

$$y(n) = \mathbf{X}^{\mathrm{T}}(n)\mathbf{W}(n), \quad (9)$$

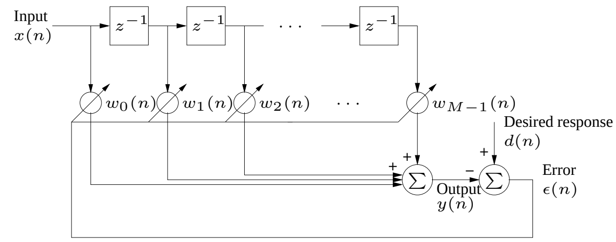

Figure 1: *Adaptive linear combiner in the form of single-input adaptive transversal filter (adapted from [7]).*

where $\mathbf{X}(n)$ contains the M linear terms of the input signal $x(n)$, the second and third order nonlinear terms generated from $x(n)$. $\mathbf{X}(n)$ can be obtained by defining the following data vectors [8]:

$$\mathbf{x}_l(n) = [x(n), x(n-1), \cdots, x(n-M+1)]^{\mathrm{T}}, \quad (10)$$

$$\mathbf{x}_2(n) = [x^2(n), x^2(n-1), \cdots, x^2(n-M+1)]^{\mathrm{T}}, \quad (11)$$

$$\mathbf{x}_3(n) = [x^3(n), x^3(n-1), \cdots, x^3(n-M+1)]^{\mathrm{T}}, \quad (12)$$

$$\mathbf{x}_{2c}(n) = [x(n)x(n-1), \cdots, x(n-M+2)x(n-M+1)]^{\mathrm{T}}, \quad (13)$$

$$\begin{aligned} \mathbf{x}_{3c}(n) = [&x(n)x(n-1)x(n-2), \\ &\cdots, x(n-M+3)x(n-M+2)x(n-M+1)]^{\mathrm{T}}, \end{aligned} \quad (14)$$

$$\begin{aligned} \mathbf{x}_{sqc}(n) = [&x^2(n)x(n-1), x^2(n-1)x(n-2), \\ &\cdots, x^2(n-M+2)x(n-M+1)]^{\mathrm{T}}, \end{aligned} \quad (15)$$

and the input vector $\mathbf{X}(n)$ in Equation (9) can be represented as

$$\mathbf{X}(n) = [\mathbf{x}_l^{\mathrm{T}}(n)\ \mathbf{x}_3^{\mathrm{T}}(n)\ \mathbf{x}_{sqc}^{\mathrm{T}}(n)\ \mathbf{x}_2^{\mathrm{T}}(n)\ \mathbf{x}_{2c}^{\mathrm{T}}(n)\ \mathbf{x}_{3c}^{\mathrm{T}}(n)]. \quad (16)$$

3.2. Obtaining Desired Response

In the adaptive process described in Equations (7) and (8), the desired response d is needed to derive the error signal ϵ and thus to update the weight vector $\mathbf{W}$ of the filter. The desired response in the adaptive system is usually obtained by a learning process or derived from the input signal, both of which are not easy in our case. When the input is composed of two sinusoidal tones, however, we can predict the output of the cubic Volterra filter; *i.e.*, the filtered output would be composed of the original two sinusoidal components and many other combination tones. Therefore, if we choose carefully the desired response, the error signal or the residual signal may consist of the CDT with the amplitude behavior seen in the experiments. We have chosen the desired response using the following processes.

First, we assumed that the amplitudes and the frequencies of the two input primary tones are known to the system. This assumption is fair enough since it is not difficult to estimate them using the frequency analysis technique such as the STFT. Second, we created a third primary tone that produces with the lower primary tone a quadratic difference tone whose frequency is the same as that of the CDT, *i.e.*, $2f_1 - f_2$. The frequency of the QDT that satisfies such a condition turns out to be $f_3 = 3f_1 - f_2$. We used the third primary tone as well as the original two primary tones as our input signals. Finally, as the desired response, we used three primary tones - the original two primary tones plus the intentionally generated third one - and the quadratic difference tone, whose amplitude behavior is known to follow closely the classical square-law behavior [2, 3]. Following this setup, we could imagine that the error or residual signal would consist of the QDT overlapped with the CDT at the same frequency of $2f_1 - f_2$ and other combination tones, which can be neglected. Figure 2 shows the block diagram of the filter.

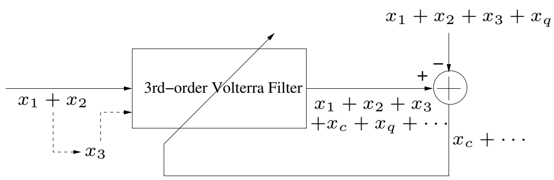

Figure 2: *Third- order adaptive Volterra filter with desired response and error signal. x_q and x_c indicate the quadratic and cubic difference tone at frequency $2f_1 - f_2$, respectively.*

The most critical point in the system is how to determine the amplitude of the QDT which will be contained in the desired response, and has a great effect on the output level of the CDT. Because our assumption was that the level of the QDT follows quite precisely the classical square-law represented by $L_{QDT} = L_1 + L_3 - C$ dB, where C depends on the relative amplitude of the quadratic distortion, and approximately is a constant as shown in [2], we could determine L_{QDT} with suitable values of L_3xs and C. In our algorithm, we set the level of the third primary tone to be the same as that of the first primary tone at frequency f_1. Ideally, the level of the QDT can be obtained using the above equation with an appropriate value of C, but it has turned out that it also requires the level of the second primary tone L_2 in order to precisely estimate the level of the CDT.

After the filter has converged, the error signal would contain not only the CDT but also other combination tones produced by three primary tones, but we concentrated on and analyzed just the CDT, since the amplitude behavior is what we are trying to explain using the filter.

4. RESULTS AND DISCUSSION

Using the third-order adaptive Volterra filter with the desired response obtained by the processes described above, we tested the algorithm with the various input signals with different input levels and frequencies of primary tones, and compared the results with the measured data provided in [2]. Figure 3 shows comparison between the experimental data and the estimated data using the filter.

In Figure 3, we can clearly see that the filtered output follows very closely the amplitude behavior of the CDT seen in the experimental data. Particularly, the decrease in the CDT with increasing levels of the lower primary can be observed in the filtered output as well, which is the largest difference between the data expected from the regular cubic distortion and the data obtained from the experiments.

Figure 4 also shows the filtered output is very close to the measured data when the frequency separation between the two primary tones increases.

Figure 5 shows the amplitude behavior of the CDT when the frequency separation is large ($\Delta f = 572$Hz). Different from the two previous figures, we can see there is an offset between the

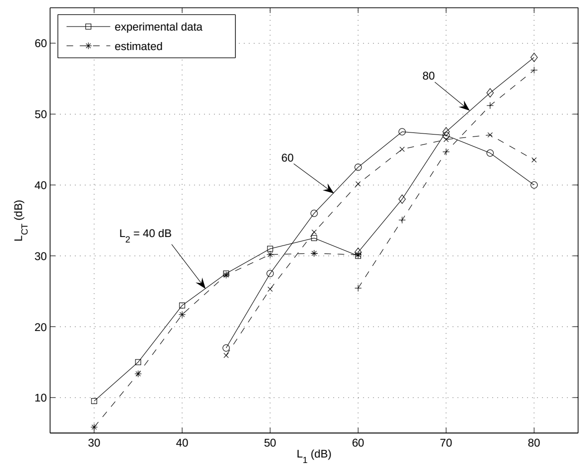

Figure 3: *Comparison between experimental data and filtered output showing the amplitude of the CDT as the input level of the lower primary varies with the higher primary level as a parameter ($f_1 = 1620$Hz, $f_2 = 1800$Hz, $\Delta f = 180$Hz). Experimental data are adapted from [2].*

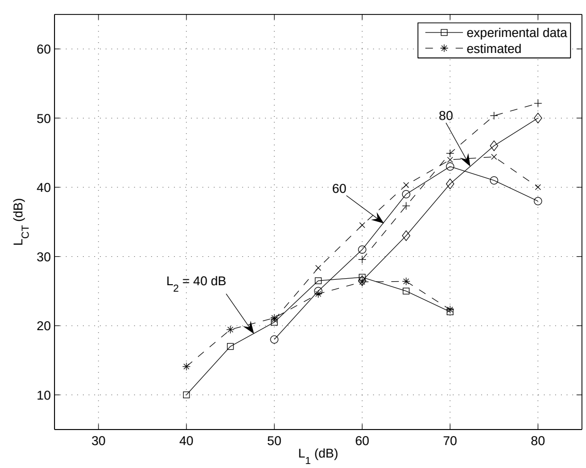

Figure 4: *Comparison between experimental data and filtered output showing the amplitude of the CDT as the input level of the lower primary varies with the higher primary level as a parameter ($f_1 = 1620$Hz, $f_2 = 1944$Hz, $\Delta f = 324$Hz). Experimental data are adapted from [2].*

measured data and the filtered output. However, the offset is quite constant at all levels, and thus can be easily compensated by controlling the level of the third primary tone we added to an input signal. In addition, the measured data shown in the figures are the means from six subjects, and the deviation is the largest in the third experiment, which is as large as about 25 dB, whereas the largest gap between the measured data and the filtered output shown in Figure 5 is less than 20 dB.

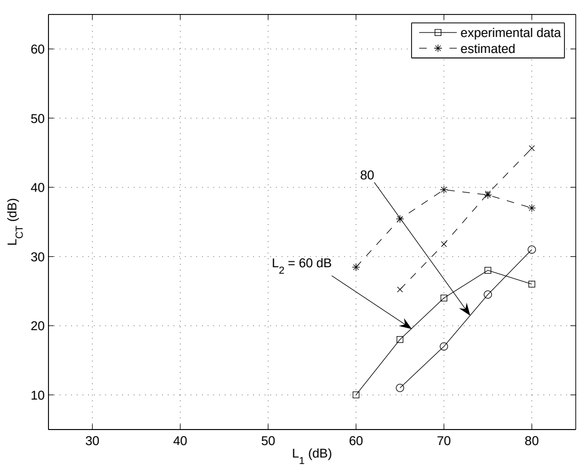

Figure 5: *Comparison between experimental data and filtered output showing the amplitude of the CDT as the input level of the lower primary varies with the higher primary level as a parameter ($f_1 = 1620$Hz, $f_2 = 2192$Hz, $\Delta f = 572$Hz). Experimental data are adapted from [2].*

As has been mentioned above, in calculating the amplitude of the quadratic difference tone to be added to the desired response, not only the amplitudes of the first and third primary tones but also that of the second primary tone was needed, which was not expected if we just think of the square-law behavior of the QDT produced by the first and third primary tones. This may be due to much more complicated nonlinear interaction between the QDT and the CDT when their frequencies coincide.

5. CONCLUSIONS

A novel model of the nonlinear amplitude behavior of the cubic difference tone has been presented using a third-order adaptive Volterra filter. In obtaining the desired response of the filter to derive the error signal, a third sinusoidal primary tone was intentionally generated and added to an input in such a way that the frequency of the quadratic difference tone produced by the first and third primary tones should coincide with that of the cubic difference tone produced by the original two primary tones. The level of the QDT was calculated using the classical square-law, and the QDT as well as three primary tones were added to the desired response. The error signal or the residual signal after convergence contained the CDT whose amplitude follows very closely what was seen in most of the experimental data provided in [2] with different intensity levels and frequencies of primary tones especially when the frequency separation is small.

6. ACKNOWLEDGMENT

The authors would like to thank Julius O. Smith for fruitful discussions and suggestions regarding this research.

7. REFERENCES

[1] H. L. F. von Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, Dover, New York, 1954, English translation of 1863 (German) edition by A. J. Ellis.

[2] E. Zwicker and H. Fastl, *Psychoacoustics*, Dover, New York, 1954, English translation of 1863 (German) edition by A. J. Ellis.

[3] J. L. Goldstein, “Auditory nonlinearity,” *Journal of the Acoustical Society of America*, vol. 41, no. 3, pp. 676–689, 1967.

[4] G. F. Smoorenburg, “Audibility region of combination tones,” *Journal of the Acoustical Society of America*, vol. 52, no. 2, pp. 603–614, 1972.

[5] H. Fletcher, *Speech and Hearing*, D. Van Nostrand Company, Inc., 1929.

[6] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, John Wiley & Sons, Inc., 2000.

[7] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice Hall, 1985.

[8] X. Li, W. K. Jenkins, and C. W. Therrien, “Computationally efficient algorithms for third order adaptive volterra filters,” *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing,*, vol. 3, pp. 1405–1408, 1998.