

**Volterra model of the parametric array loudspeaker operating at ultrasonic
frequencies**

Chuang Shi^{a)} and Yoshinobu Kajikawa

Department of Electrical and Electronic Engineering

Kansai University

3-3-35 Yamatecho, Suita, Osaka Prefecture, Japan 564-8680

^{a)}e-mail: r148005@kansai-u.ac.jp

Abstract

The parametric array loudspeaker (PAL) is an application of the parametric acoustic array in air, which can be applied to transmit a narrow audio beam from an ultrasonic emitter. However, nonlinear distortion is very perceptible in the audio beam. Modulation methods to reduce the nonlinear distortion are available for on-axis far-field applications. For other applications, preprocessing techniques are wanting. In order to develop a preprocessing technique with general applicability to a wide range of operating conditions, the Volterra filter is investigated as a nonlinear model of the PAL in this paper. Limitations of the standard audio-to-audio Volterra filter are elaborated. An improved ultrasound-to-ultrasound Volterra filter is proposed and empirically demonstrated to be a more generic Volterra model of the PAL.

©2016 Acoustical Society of America

PACS numbers: 43.25.Lj, 43.60.Mn

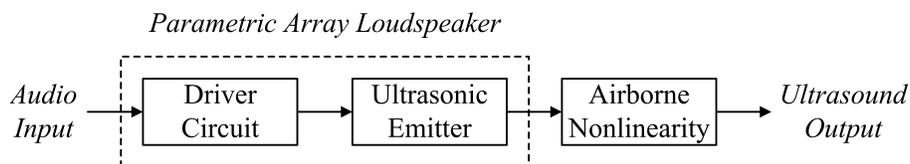


Figure 1: Block diagram of the parametric array loudspeaker.

I. INTRODUCTION

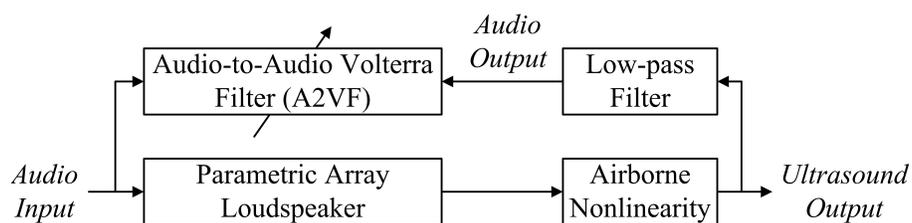
When two large-amplitude waves at close frequencies propagate in the same direction, virtual sources of the difference frequency are created, forming an end-fire array. This nonlinear acoustic phenomena was discovered by Westervelt for underwater applications and named the parametric acoustic array (PAA)¹. The PAA in air was subsequently demonstrated by Bennett and Blackstock². The first directional sound device making use of the PAA in air was established in 1983, which is now widely known as the parametric array loudspeaker (PAL)³.

The common design of the PAL, as shown in Fig. 1, consists of a driver circuit and an ultrasonic emitter. The driver circuit carries out modulation and amplification. The modulated audio input falls in the ultrasonic frequency band. It is referred to as the ultrasound input in the latter part of this paper. After the ultrasound input is transmitted from the ultrasonic emitter, the PAA is formed in air and results in a narrower audio beam than other sound devices of the same size.

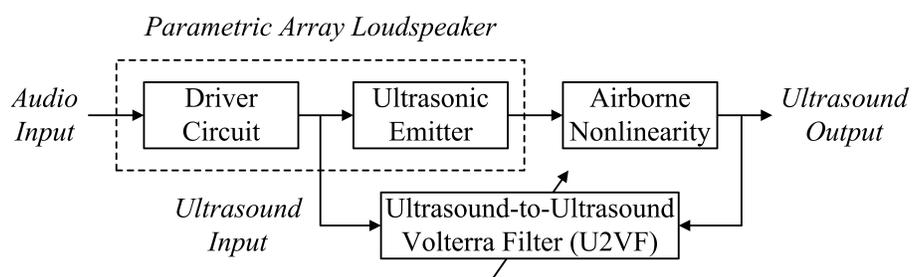
Nonlinear distortion is an adverse by-product of the PAA in air. Modulation methods have been proposed to reduce the nonlinear distortion^{4;5;6;7}. Audio bandwidth extension techniques have also been tried to improve the perceptual sound quality of the PAL^{8;9}. All those methods were developed on the basis of the Berktaf equation that formulated the principle of the PAL concisely^{10;11}. However, the far-field and absorption-limited conditions assumed by Berktaf are not always valid in applications of the PAL, such as in home entertainment¹², private listening¹³, noise canceling devices^{14;15}, and so forth.

To develop a preprocessing technique with general applicability to a wide range of operating conditions, a generic nonlinear model of the PAL is necessary. Although the second order nonlinear acoustic equation serves this purpose, it has only numerical solutions that lead to huge computational burdens¹⁶. The Volterra filter, in comparison, can be conveniently identified in audio systems^{17;18;19}. A dedicated inverse system is then designed to preprocess the audio input in order to compensate for the nonlinear distortion²⁰. Therefore, the standard audio-to-audio Volterra filter (A2VF) has been used to model the PAL from the audio input to the audio output, as illustrated in Fig. 2(a). However, it may be overlooked that the A2VF is not optimum, because the PAL is a nonlinear system operating at ultrasonic frequencies and the nonlinearity is not incurred in the PAL from the audio input to the audio output.

In this paper, the ultrasound-to-ultrasound Volterra filter (U2VF) is suggested as an



(a) Audio-to-audio Volterra filter (A2VF)



(b) Ultrasound-to-ultrasound Volterra filter (U2VF)

Figure 2: Block diagram of the (a) audio-to-audio and (b) ultrasound-to-ultrasound Volterra filters, where the slanted arrow indicates an adaptive mechanism to obtain coefficients of the Volterra filters.

improved Volterra model of the PAL. The U2VF models the PAL from the ultrasound input to the ultrasound output, as illustrated in Fig. 2(b). In Sections II and III, it will be shown that when the Berktay equation is extended to include the frequency response of the ultrasonic emitter, it is readily carried out by the parallel cascade structure of the U2VF. In Section IV, experimental results are presented to demonstrate that the U2VF is a more generic Volterra model than the A2VF, when there are changes in the modulation method and modulation index.

II. THEORY

A. Berktay Equation

As mentioned, the Berktay equation is the basis of modulation methods of the PAL.

The primary source strength density q_1 is given by

$$q_1 = \frac{\beta P_0^2}{\rho_0^2 c_0^4} \exp(-2\alpha_0 z) \frac{\partial}{\partial t} [E(t) \cos(\omega_c t)]^2, \quad (1)$$

where β is the nonlinear coefficient; ρ_0 is the density of air; and c_0 is the speed of sound in air; P_0 , α_0 , and ω_c are initial amplitude, attenuation rate, angular frequency of the ultrasonic carrier, respectively; z is the on-axis coordinate; t is the retarded time; and $E(t)$ is the envelope function, which varies slowly compared to the ultrasonic carrier¹⁰.

The self-demodulated source strength density is extracted from (1), keeping only the

audible frequency component as

$$q_d = \frac{\beta P_0^2}{\rho_0^2 c_0^4} \exp(-2\alpha_0 z) \frac{\partial}{\partial t} \left[\frac{E^2(t)}{2} \right]. \quad (2)$$

The self-demodulated pressure is calculated by

$$p_d = \frac{\rho_0 S_0}{4\pi} \int_0^{z'} \frac{1}{z' - z} \frac{\partial q_d}{\partial t} dz = \frac{\beta P_0^2 S_0}{8\pi \rho_0 c_0^4} u(l_a) \frac{\partial^2}{\partial t^2} E^2(t), \quad (3)$$

where z' is the observation point; S_0 is the source size; l_a is the effective length of the PAA;

and it is also defined for simplicity that

$$u(l_a) = \int_0^{l_a} \frac{\exp(-2\alpha_0 z)}{z' - z} dz. \quad (4)$$

When the far-field and absorption-limited conditions are assumed, substituting

$u(+\infty) = 1/2\alpha_0 z'$ into (3) yields the Berkay equation as

$$p_d = \frac{\beta P_0^2 S_0}{16\pi \rho_0 c_0^4 \alpha_0 z'} \frac{\partial^2}{\partial t^2} E^2(t). \quad (5)$$

B. Modulation Methods

In the first known PAL, the double sideband (DSB) modulation method was adopted³. The envelope function of the DSB modulation method is written as

$$E_{DSB}(t) = 1 + mA(t), \quad (6)$$

where m is the modulation index and $A(t)$ is the audio input. The frequency response of the PAL decreases with a slope of 12 dB per octave as a result of the second derivative in

(5). Moreover, the second harmonic ratio (SHR), *i.e.* the ratio of the amplitude of the second harmonic to that of the fundamental frequency, is proportional to the modulation index.

As a direct inverse system to the Berkta equation, the square root (SRT) modulation method was introduced to eliminate the nonlinear distortion of the PAL⁴. The envelope function of the SRT modulation method is written as

$$E_{SRT}(t) = \sqrt{1 + mA(t)}. \quad (7)$$

The drawback of the SRT modulation method occurs in the implementation stage, because the ultrasonic emitter has neither an infinite bandwidth nor a flat frequency response⁵.

Therefore, the Taylor expansion of (7) is more meaningful in the performance analysis of the SRT modulation method, which is written as

$$E_{SRT}(t) = 1 + \frac{m}{2}A(t) - \frac{m^2}{8}A^2(t) + O(A^3), \quad (8)$$

where $O(A^3)$ denotes the third and higher order terms. When the modulation index is small, the SRT modulation method can be approximated by the DSB modulation method using half the modulation index.

C. Extended Berkta Equation

The ultrasonic emitter of the PAL is made up of a number of piezoelectric transducers

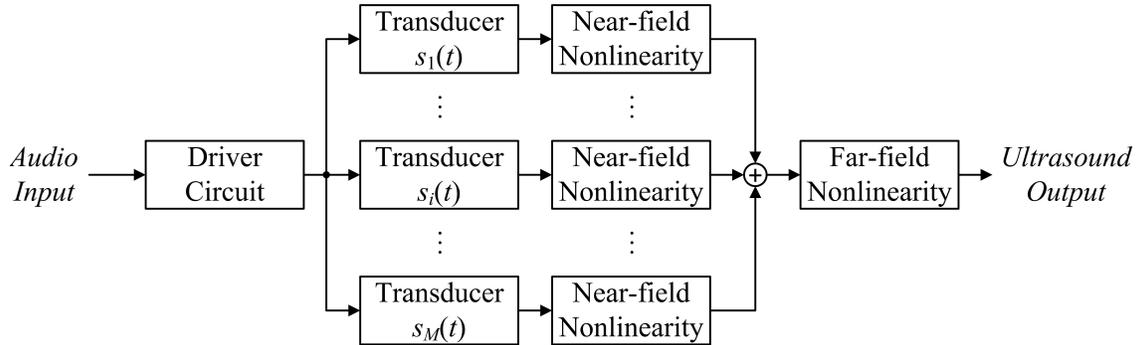


Figure 3: Block diagram of the parametric array loudspeaker assuming two stages of self-demodulation.

(PZTs) that have individual frequency responses²¹. In order to take them into account, the extended Berklay equation is derived inductively as follows.

Fig. 3 illustrates a tentatively assumed case where there are two subsequent stages of self-demodulation along the axis of the ultrasonic emitter. The first stage takes place in the near-field of individual PZTs. The second stage takes place in the far-field of individual PZTs but within the near-field of the whole ultrasonic emitter.

In the near-field of individual PZTs, they are treated as piston sources. Primary waves transmitted by a PZT do not overlap with those transmitted by other PZTs. Therefore, self-demodulated waves are independently generated. If the impulse response of the i th PZT is denoted as $s_i(t)$, the superposition of self-demodulated waves yields the

audio output of the PAL as

$$p_n = \frac{\beta P_0^2 S_0}{8\pi\rho_0 c_0^4 M} u(l_n) \sum_{i=1}^M \{s_a(t) * \frac{\partial^2}{\partial t^2} \{s_i(t) * [E(t) \cos(\omega_c t)]\}^2\}, \quad (9)$$

where M is the total number of PZTs; l_n denotes the range of the near-field of individual PZTs; $s_a(t)$ is the impulse response of a low-pass filter to suppress ultrasonic frequencies, modeling the extraction of (2) from (1); and $*$ denotes linear convolution. Because the second derivative and convolution are linear operators, the sequence of them is exchangeable.

Beyond the near-field of individual PZTs, primary waves transmitted by neighboring PZTs start to overlap and result in large virtual sources. It is impractical to measure acoustically the characteristics of virtual sources. To simplify the derivation, we have assumed that virtual sources are created in the same size as the ultrasonic emitter immediately after the first stage of self-demodulation. Therefore, the audio output of the PAL contributed by the second stage of self-demodulation is expressed by

$$p_f = \frac{\beta P_0^2 S_0}{8\pi\rho_0 c_0^4} [u(l_a) - u(l_n)] \{s_a(t) * \frac{\partial^2}{\partial t^2} \{ \sum_{i=1}^M s_i(t) * [E(t) \cos(\omega_c t)] \}^2\}. \quad (10)$$

Eqs. (9) and (10) can be summarized into a general formula that is not limited to two stages of self-demodulation as

$$p_d = s_a(t) * \sum_{i=1}^{\hat{M}} K_i \{ \hat{s}_i(t) * [E(t) \cos(\omega_c t)] \}^2, \quad (11)$$

where K_i is a positive value, related to the effective length of a group of PZTs; $\hat{s}_i(t)$ is the overall frequency response of the group of PZTs, combined with the second derivative; and the upper bound of the summation is denoted by \hat{M} , depending on the number of PZTs and stages of self-demodulation. For instance, we have tentatively assumed in Fig. 3 that

$$p_d(x) = p_n(x) + p_f(x), \quad (12)$$

whereby $\hat{M} = M + 1$; $\hat{s}_i(t) = s_i(t)$ and $K_i = \beta P_0^2 S_0 u(l_n) / 8\pi \rho_0 c_0^4 M$, when $i \leq M$; $\hat{s}_{M+1}(t) = \sum_{i=1}^M s_i(t)$ and $K_{M+1} = \beta P_0^2 S_0 [u(l_a) - u(l_n)] / 8\pi \rho_0 c_0^4$.

Eq. (11) is referred to as the extended Berktay equation in this paper. It will be elaborated in the next section that (11) is readily carried out by the parallel cascade structure of the U2VF. When PZTs are assumed to have the same frequency response, the extended Berktay equation is simplified to the Berktay equation with an overall frequency response of the ultrasonic emitter. This special case has been examined with the one-dimension A2VF previously²².

III. MODELING METHOD

A. Volterra Filter

Nonlinearity of the PAL can be modeled by the Volterra filter as

$$y_n = \sum_{i=0}^{N-1} h_1(i) x_{n-i} + \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} h_2(i, j) x_{n-i} x_{n-j} + \varepsilon_n, \quad (13)$$

where x_n and y_n are the input and output samples at the discrete time index n ; ε_n is the model error, containing the third and higher order nonlinearity; N is the memory length; h_1 and h_2 are the first and second order Volterra kernels, respectively²³.

Eq. (13) can be rewritten in the vector notation as

$$y_n = \mathbf{H}^T \mathbf{X}_n + \varepsilon_n, \quad (14)$$

where

$$\begin{aligned} \mathbf{H} = & [h_1(0), h_1(1), \dots, h_1(N-1), \\ & h_2(0,0), h_2(0,1), \dots, h_2(N-1, N-1)]^T \end{aligned} \quad (15)$$

and

$$\mathbf{X}_n = [x_n, x_{n-1}, \dots, x_{n-N+1}, x_n x_n, x_n x_{n-1}, \dots, x_{n-N+1} x_{n-N+1}]^T. \quad (16)$$

Both \mathbf{H} and \mathbf{X}_n are vectors with the length of $N^2 + N$. The normalized least mean squares (NLMS) algorithm can be adopted to estimate \mathbf{H} iteratively as

$$\mathbf{H}_{n+1} = \mathbf{H}_n + \mu \frac{\mathbf{X}_n \varepsilon_n}{\mathbf{X}_n^T \mathbf{X}_n}, \quad (17)$$

where μ is the step size^{24;25}.

B. Parallel Cascade Structure

Computational complexity of the Volterra filter can be reduced by adopting the parallel cascade structure^{26;27}. When h_2 is real symmetric, it is decomposed into

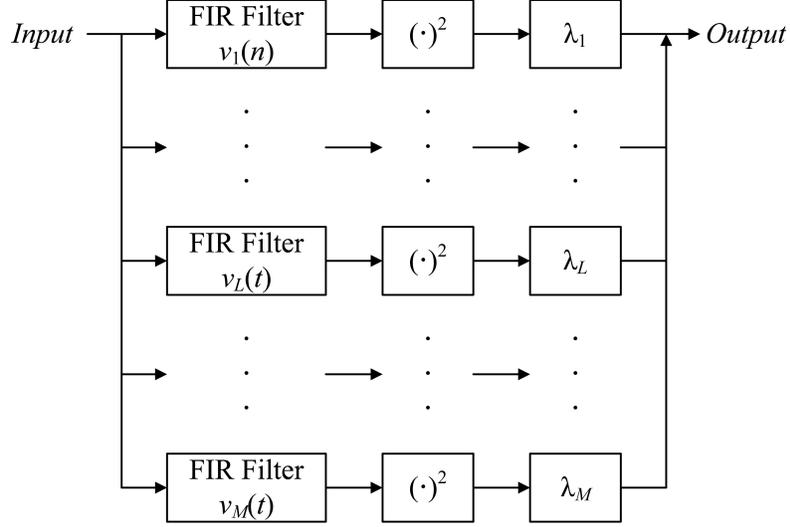


Figure 4: Parallel cascade structure of the second order Volterra kernel, where $v_i(n)$ denotes an eigenvector of the second order Volterra kernel and λ_i denotes the corresponding eigenvalues.

$h_2 = V^T \Lambda V$, where $\Lambda := \text{diag}(\lambda_1, \dots, \lambda_N)$ is a diagonal matrix consisting of eigenvalues of h_2 and rows of V are the corresponding eigenvectors. Providing the input vector $\mathbf{x} = [x_n, x_{n-1}, \dots, x_{n-N+1}]^T$, the second order nonlinear output is manipulated as

$$\mathbf{x}^T h_2 \mathbf{x} = \mathbf{x}^T (V^T \Lambda V) \mathbf{x} = (V \mathbf{x})^T \Lambda (V \mathbf{x}). \quad (18)$$

Eq. (18) provides the basis of the parallel cascade structure. The i th row of V is treated as a finite impulse response (FIR) filter, of which the coefficients are denoted as $v_i(n)$ for convenience. Each element of $V \mathbf{x}$ equals the filtered output of \mathbf{x} . Therefore, (18)

is interpreted as a linear combination of the squared outputs of a bank of FIR filters, as illustrated in Fig. 4. When there are $M - L$ trivial eigenvalues of h_2 , keeping only the L largest eigenvalues saves the computational cost of implementing the Volterra filter.

When $x(n) = E(n) \cos(\omega_c n)$, the audio output of the PAL calculated by the parallel cascade structure of the U2VF is written as

$$p_d(n) = s_a(n) * \sum_{i=1}^L \lambda_i \{v_i(n) * [E(n) \cos(\omega_c n)]\}^2. \quad (19)$$

This is an equivalent form of the extended Berktay equation. Eigenvalue λ_i and eigenvector $v_i(n)$ correspond to parameter K_i and impulse response $\hat{s}_i(t)$, respectively. When PZTs in the ultrasonic emitter have similar frequency responses, L can be much smaller than M . Particularly when $L = 1$, (19) is simplified to the one-dimension Volterra filter, where only the main diagonal elements are implemented for the second order Volterra kernel²². By doing so, the computational cost is greatly reduced but the model accuracy is traded off.

IV. Experimental Validation

The experiment is carried out in a sound proof room ($2.9 \times 3.1 \times 2.1 \text{ m}^3$), where the microphone (B&K 4191L) is placed 3.0 m away from the ultrasonic emitter (Mitsubishi MSP-30E). The microphone is a half-inch free-field microphone with a designed frequency range from 3 Hz to 40 kHz. The ultrasonic emitter is made up of more than 200 PZTs and has a diameter of 18 cm. Both the digital-to-analog and analog-to-digital converters have

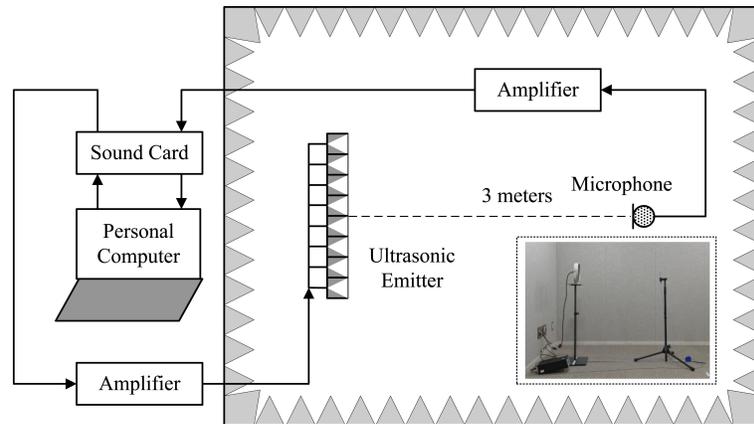


Figure 5: (color online) Experimental setup for evaluating the Volterra filters that model the nonlinearity of the parametric array loudspeaker, with an inserted photograph.

the sampling frequency of 192 kHz and the resolution of 32 bit. The experimental setup is illustrated in Fig. 5.

The carrier frequency is chosen at 40 kHz. A band-passed white noise from 32 kHz to 48 kHz is transmitted from the ultrasonic emitter for the U2VF identification. The ultrasound level is maintained at 110 dB. Thereafter, the DSB and SRT modulation methods are implemented with the modulation index m changing from 0.1 to 1.0 with an interval of 0.1. A group of A2VFs are identified for the DSB modulation method with different modulation indexes by the same low-passed white noise cut off at 8 kHz. A2VFs are not identified for the SRT modulation method, because the SRT modulation method leads to infinite harmonics of the fundamental frequency. Since the Volterra filter is

truncated at the second order in this paper, it is inadequate to model the PAL adopting the SRT modulation method. This is an obvious limitation of the A2VF when the modulation method introduces high order nonlinearity.

The sound speed in dry air at 20°C is estimated to be 343 m/s. The acoustic delay from the ultrasonic emitter to the microphone is 8.75 ms, which equals 1680 samples at the sampling frequency of 192 kHz. The causality of the Volterra filter is not satisfied if the memory length is set shorter than the acoustic delay. On the other hand, because the computational complexity increases exponentially with the memory length, using a long memory length is not desired. The input and output of the identification have to be aligned so that the memory length can be shortened. In this paper, the acoustic delay is offset by 1600 samples and the memory length is set to 300 samples.

Total harmonic distortion (THD) and intermodulation distortion (IMD) are two widely used performance measures of the PAL. The THD level is defined as the ratio of the root mean square (RMS) amplitude of the audible harmonics to the RMS amplitude of the fundamental frequency. The IMD level is defined as the ratio of the RMS amplitude of the audible intermodulation frequencies to the RMS amplitude of the two fundamental frequencies. The identified U2VF and A2VFs are evaluated by the simulated THD and IMD tests. The THD and IMD levels are also measured by the experimental setup in Fig. 5. Model accuracies of the U2VF and A2VFs are judged by how close the simulation

results can match the measurement results. A sine sweep is used as the testing audio input, of which the frequency varies from 0.5 kHz to 8 kHz. In the IMD test, another sine tone at 1.7 kHz is added in. The amplitude of the sine tone is four times as that of the sine sweep.

THD curves of the DSB and SRT modulation methods are plotted in Fig. 6. In Figs. 6(a) and (b), it is shown that the A2VF identified for a specific modulation index can only be used to predict the THD level of the same modulation index. To make the comparison between the U2VF and A2VFs for the SRT modulation method, the A2VFs identified for the DSB modulation method are used with the modulation index halved. It is shown in Fig. 6(c) that the A2VF identified for $m = 0.25$ predicts the THD level of the SRT modulation method when $m = 0.5$ more accurately than the A2VF identified for $m = 0.5$. In Fig. 6(d), since the modulation index is large, the approximation implied by (8) is no longer accurate. Therefore, the A2VF identified for $m = 0.5$ is only adequate to predict the THD level of the SRT modulation method when the testing frequency is above 5 kHz. In such a case, the frequency response of the ultrasonic emitter works like a band-pass filter so that the PAL cannot generate the third and higher order harmonics that are audible. When the testing frequency is below 5 kHz, the A2VF identified for $m = 0.5$ tends to under estimate the THD level of the SRT modulation, because the third and higher order harmonics become audible and they are not modeled by A2VFs truncated at the second order.

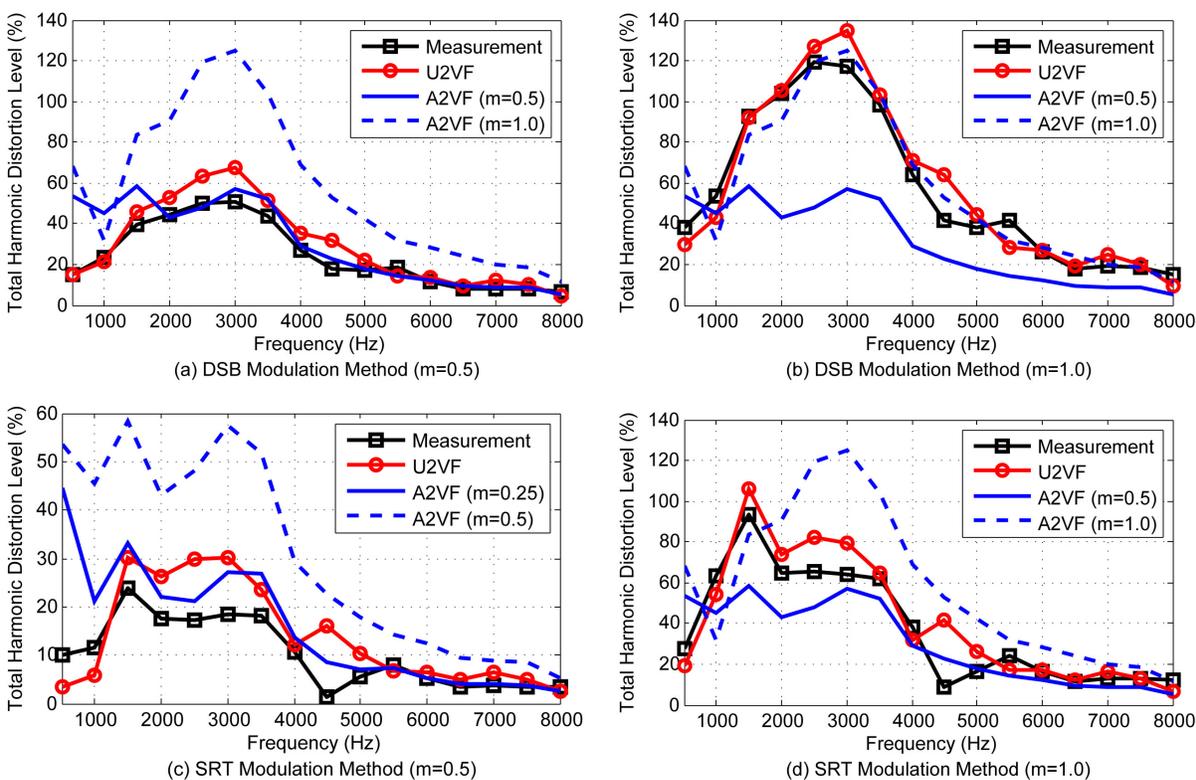


Figure 6: (color online) Total harmonic distortion curves of (a) DSB modulation method using the modulation index of 0.5; (b) DSB modulation method using the modulation index of 1.0; (c) SRT modulation method using the modulation index of 0.5; (d) SRT modulation method using the modulation index of 1.0.

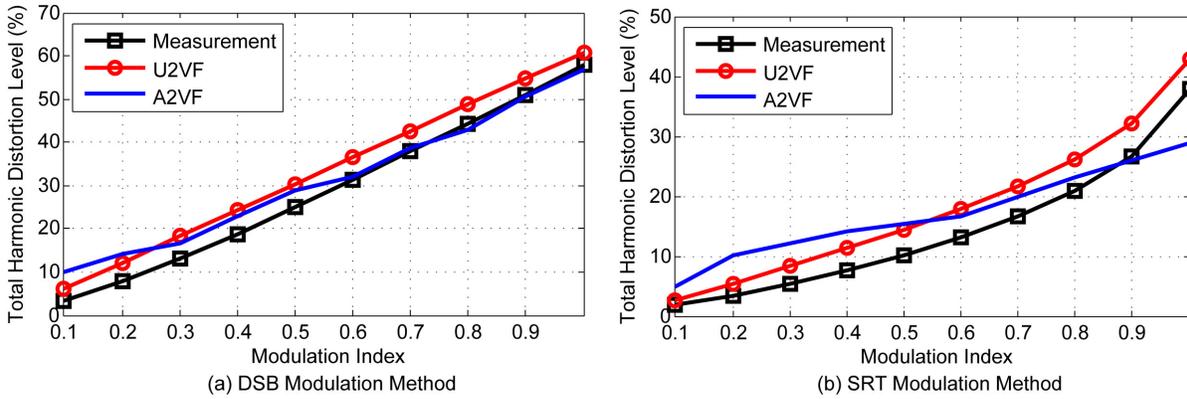


Figure 7: (color online) Averaged total harmonic distortion curves of (a) DSB modulation method; (b) SRT modulation method.

In Fig. 7, the averaged THD levels across the testing frequencies are plotted with respect to the modulation index. Fig. 7 shows that when there are changes in the modulation method and modulation index, the U2VF can always predict the THD level of the PAL with certain accuracy. The model accuracy of the A2VF is only high when the identification and THD measurement takes place under identical conditions. The model accuracy of the A2VF is found to decrease with the modulation index. This is because when the modulation index is low, the second order nonlinearity becomes weak in the audible frequency band. In this case, the output of the first order Volterra kernel is about 20 dB higher than the output of the second order Volterra kernel. The second harmonic amplitude may be lower than the noise floor. The U2VF overcomes this problem because

the second order nonlinear output remains strong in the ultrasonic frequency band.

Therefore, using a microphone that can measure the second harmonic of the ultrasonic input benefits the model accuracy of the U2VF. In this paper, due to the limited frequency range of the microphone, the model accuracy of the U2VF is slightly compromised.

Fig. 8 shows the frequency responses of the DSB and SRT modulation methods when $m = 0.5$ and $m = 1.0$. General trends are similar, since they are obtained with the same ultrasonic emitter. When the modulation index is specified, the sound pressure level of the SRT modulation method is lower than that of the DSB modulation method. When the modulation method is fixed, higher modulation index leads to higher sound pressure level.

Furthermore, IMD curves and averaged IMD curves are plotted in Figs. 9 and 10, respectively. It is noteworthy that when the A2VFs identified for the DSB modulation method are used for the SRT modulation method, the modulation indexes are halved. Similar to the observations obtained in Figs. 6 and 7, the U2VF provides a more generic Volterra model of the PAL, whereas the A2VF lacks the flexibility upon changes in the modulation method and modulation index.

Changing the modulation method is not a common practice after the PAL is deployed in an application. However, when designing the PAL, we need to evaluate the performance of different modulation methods. The A2VF is shown to be inadequate to serve this

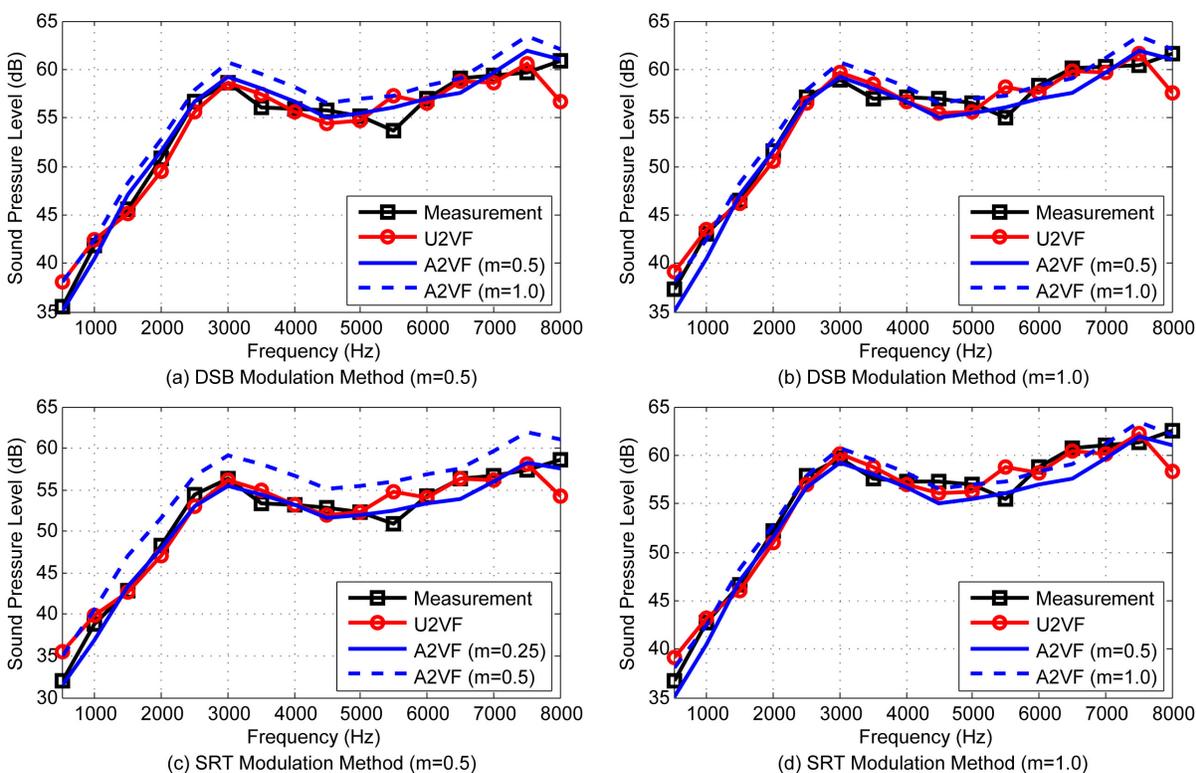


Figure 8: (color online) Frequency responses of (a) DSB modulation method using the modulation index of 0.5; (b) DSB modulation method using the modulation index of 1.0; (c) SRT modulation method when using the modulation index of 0.5; (d) SRT modulation method using the modulation index of 1.0.

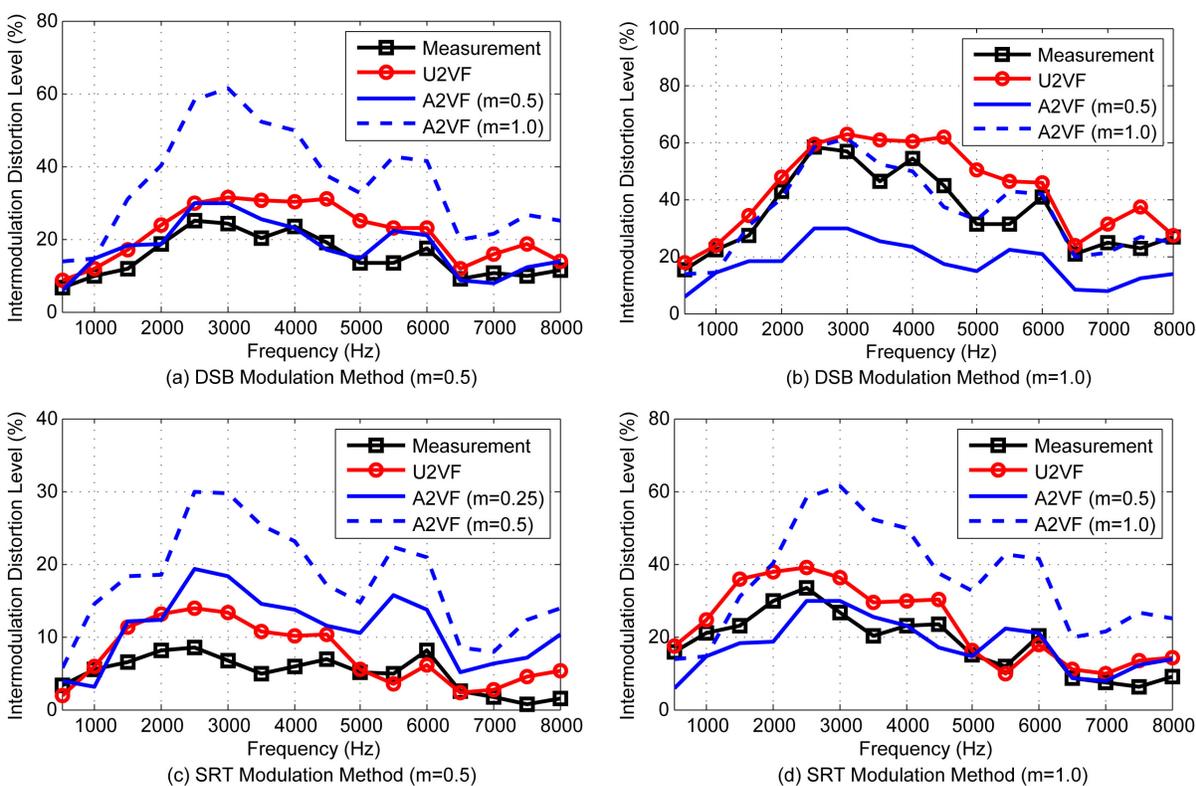


Figure 9: (color online) Intermodulation distortion curves of (a) DSB modulation method using the modulation index of 0.5; (b) DSB modulation method using the modulation index of 1.0; (c) SRT modulation method using the modulation index of 0.5; (d) SRT modulation method using the modulation index of 1.0.

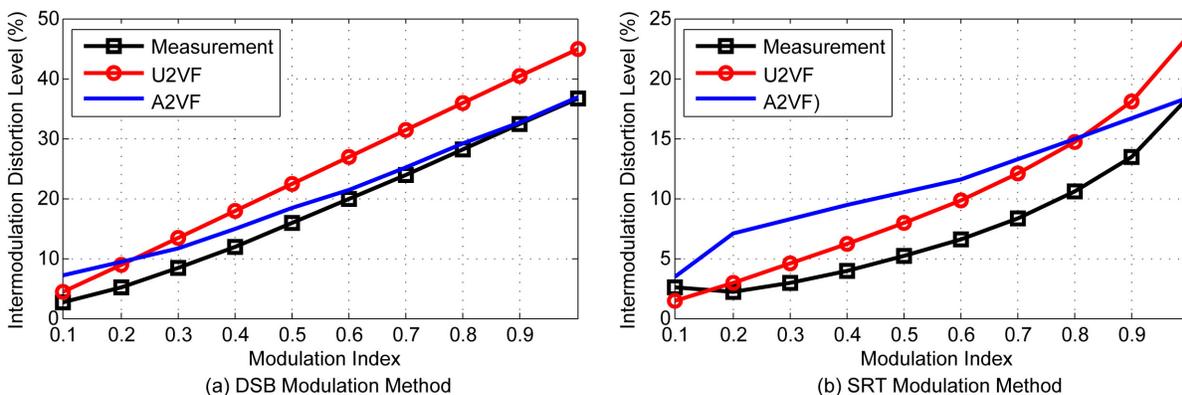


Figure 10: (color online) Averaged intermodulation distortion curves of (a) DSB modulation method; (b) SRT modulation method.

purpose, because many modulation methods introduce high order nonlinearity. Moreover, changes in the modulation index are more often incurred in the PAL. Since the modulation index is multiplied with the audio input, time varying amplitude of the audio input is equivalent to time varying modulation index.

V. Conclusions

Two types of Volterra filters have been studied to model the nonlinearity of the PAL in order that preprocessing techniques can be developed and evaluated under a wide range of operating conditions. The A2VF, although a standard in audio systems, encounters difficulties when the modulation method and modulation index are changed. Therefore, it may not be useful during the evaluation of preprocessing techniques that introduce high

order nonlinearity. In contrast, the U2VF recommended by this paper has demonstrated constantly high model accuracy upon changes in the modulation method and modulation index. As a more generic Volterra model of the PAL, the U2VF can be used to synthesize different A2VFs, without the need to carry out extra measurements. Existing linearization methods are applicable to those synthesized A2VFs. A combination of the synthesis and linearization methods will eventually lead to the development of a sophisticated preprocessing technique that reduces the perceptible nonlinear distortion of the PAL better than the state of the art.

Acknowledgements

This work is supported by MEXT-Supported Program for the Strategic Research Foundation at Private Universities, 2013-2017.

REFERENCES

1. P. J. Westervelt, "Parametric acoustic array," *J. Acoust. Soc. Am.* **35**, 535-537 (1963).
2. M. B. Bennett and D. T. Blackstock, "Parametric array in air," *J. Acoust. Soc. Am.* **57**, 562-568 (1975).
3. M. Yoneyama, J. Fujimoto, Y. Kawamo, and S. Sasabe, "The audio spotlight: An

- application of nonlinear interaction of sound waves to a new type of loudspeaker design,” *J. Acoust. Soc. Am.* **73**, 1013–1020 (1983).
4. T. Kamakura, M. Yoneyama, and K. Ikegaya, “Developments of parametric loudspeaker for practical use,” *Proc. 10th Int. Symp. Nonlinear Acoust.*, Kobe, Japan, 147–150 (1984).
 5. K. Aoki, T. Kamakura, and Y. Kumamoto, “Parametric loudspeaker: Characteristics of acoustic field and suitable modulation of carrier ultrasound,” *Electron. Commun. Jpn.*, **74**, 76–82 (1991).
 6. T. D. Kite, J. T. Post, and M. F. Hamilton, “Parametric array in air distortion reduction by preprocessing,” *J. Acoust. Soc. Am.* **103**, 2871 (1998).
 7. F. J. Pompei, “The use of airborne ultrasonics for generating audible sound beams,” *J. Audio Eng. Soc.*, **47**, 726–731 (1999).
 8. C. Shi, H. Mu, and W. S. Gan, “A psychoacoustical preprocessing technique for virtual bass enhancement of the parametric loudspeaker,” *Proc. 38th Int. Conf. Acoust. Speech Signal Process.*, Vancouver, Canada, 31–35 (2013).
 9. C. Shi and W. S. Gan, “A preprocessing method to increase high frequency response

- of a parametric loudspeaker,” *Proc. 2013 APSIPA Annu. Summit Conf.*, Kaohsiung, Taiwan, 1–5 (2013).
10. H. O. Berktag, “Possible exploitation of non-linear acoustics in under-water transmitting applications,” *J. Sound Vib.* **2**, 435-461 (1965).
 11. H. M. Merklinger, “Improved efficiency in the parametric transmitting array,” *J. Acoust. Soc. Am.* **58**, 784–787 (1975).
 12. S. Aoki, M. Toba, and N. Tsujita, “Sound localization of stereo reproduction with parametric loudspeakers,” *Applied Acoust.*, **73**, 1289–1295 (2012).
 13. A. Wright, A. Evans, A. Linney, and M. Lincoln, “The listening room: A speech-based interactive art installation,” *Proc. 15th Int. Conf. Multimedia*, Augsburg, Germany, 1–10 (2007).
 14. N. Tanaka and M. Tanaka, “Active noise control using a steerable parametric array loudspeaker,” *J. Acoust. Soc. Am.* **127**, 3526–3537 (2010).
 15. K. Tanaka, C. Shi, and Y. Kajikawa, “Binaural active noise control using parametric array loudspeakers,” *Applied Acoust.* **116**, 170–176 (2017).
 16. Y. S. Lee and M. F. Hamilton, “Time-domain modeling of pulsed finite-amplitude sound beams,” *J. Acoust. Soc. Am.*, **97**, 906–917 (1995).

17. C. M. Lee, J. Yang, W. S. Gan, and M. H. Er, “Modeling nonlinearity of air with Volterra kernels for use in a parametric array loudspeaker,” *Proc. 112th AES Conv.*, Munich, Germany, 5635 (2002).
18. T. Helie and M. Hasler, “Volterra series for solving weakly non-linear partial differential equations: Application to a dissipative Burgers’ equation,” *Int. J. Control*, **77**, 1071–1082 (2004).
19. W. Ji and W. S. Gan, “Identification of a parametric loudspeaker system using an adaptive Volterra filter,” *Appl. Acoust.*, **73**, 1251–1262 (2012).
20. Y. Hatano, C. Shi, and Y. Kajikawa, “A study on linearization of nonlinear distortions in parametric array loudspeakers,” *Proc. 2014 Int. Workshop Smart Inf. Media Syst.*, Ho Chi Minh City, Vietnam, 79–82 (2014).
21. T. Kamakura, H. Nomura, and G. T. Clement, “Linear and nonlinear ultrasound fields formed by planar sources with random pressure distributions,” *Acoust. Sci. Technol.*, **36**, 208–215 (2015).
22. Y. Mu, P. Ji, W. Ji, M. Wu, and J. Yang, “Modeling and compensation for the distortion of parametric loudspeakers using a one-dimension Volterra filter,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **22**, 2169–2181 (2014).

23. V. J. Mathews, "Adaptive polynomial filters," *IEEE Sig. Process. Mag.*, **8**, 10–26 (1991).
24. C. Shi and Y. Kajikawa, "Identification of the parametric array loudspeaker with a Volterra filter using the sparse NLMS algorithm," *Proc. 40th IEEE Int. Conf. Acoust. Speech Signal Process.*, Brisbane, Australia, 3372–3376 (2015).
25. C. Shi and Y. Kajikawa, "Ultrasound-to-ultrasound Volterra filter identification of the parametric array loudspeaker," *Proc. 20th IEEE Int. Conf. Digital Signal Process.*, Singapore, 1–4 (2015).
26. O. Hoshuyama and T. Okayasu, "Decimation in Volterra filter for distortion reduction of ultrasonic parametric loudspeaker," *Proc. 2014 IEICE Soc. Conf.*, Tokushima, Japan, 82 (2014).
27. Y. Hatano, C. Shi, S. Kinoshita and Y. Kajikawa, "A study on linearization of nonlinear distortions in parametric array loudspeakers," *Proc. 23rd European Sig. Process. Conf.*, Nice, France, 1073-1077 (2015).

Figure Captions

Figure 1. Block diagram of the parametric array loudspeaker.

Figure 2. Block diagram of the (a) audio-to-audio and (b) ultrasound-to-ultrasound Volterra filters, where the slanted arrow indicates an adaptive mechanism to obtain coefficients of the Volterra filters.

Figure 3. Block diagram of the parametric array loudspeaker assuming two stages of self-demodulation.

Figure 4. Parallel cascade structure of the second order Volterra kernel, where $v_i(n)$ denotes an eigenvector of the second order Volterra kernel and λ_i denotes the corresponding eigenvalues.

Figure 5. (color online) Experimental setup for evaluating the Volterra filters that model the nonlinearity of the parametric array loudspeaker, with an inserted photograph.

Figure 6. (color online) Total harmonic distortion curves of (a) DSB modulation method using the modulation index of 0.5; (b) DSB modulation method using the modulation index of 1.0; (c) SRT modulation method using the modulation index of 0.5; (d) SRT modulation method using the modulation index of 1.0.

Figure 7. (color online) Averaged total harmonic distortion curves of (a) DSB modulation method; (b) SRT modulation method.

Figure 8. (color online) Frequency responses of (a) DSB modulation method using the modulation index of 0.5; (b) DSB modulation method using the modulation index of 1.0; (c) SRT modulation method when using the modulation index of 0.5; (d) SRT modulation method using the modulation index of 1.0.

Figure 9. (color online) Intermodulation distortion curves of (a) DSB modulation method using the modulation index of 0.5; (b) DSB modulation method using the modulation index of 1.0; (c) SRT modulation method using the modulation index of 0.5; (d) SRT modulation method using the modulation index of 1.0.

Figure 10. (color online) Averaged intermodulation distortion curves of (a) DSB modulation method; (b) SRT modulation method.