A comparative analysis of preprocessing methods for the parametric loudspeaker based on the Khokhlov-Zabolotskaya-Kuznetsov equation for speech reproduction

Gan, Woon-Seng; Yang, Jun.; Ji, Peifeng.; Tan, Ee-Leng.

2010

Ji, P., Tan, E. L., Gan, W. S., & Yang, J. (2010). A Comparative Analysis of Preprocessing Methods for the Parametric Loudspeaker Based on the Khokhlov-Zabolotskaya-Kuznetsov Equation for Speech Reproduction. IEEE Transactions on Audio, Speech, and Language Processing, 19(4), 937-946.

https://hdl.handle.net/10356/93869

https://doi.org/10.1109/TASL.2010.2070794

A comparative analysis of preprocessing methods for the parametric loudspeaker based on the Khokhlov-Zabolotskaya-Kuznetsov equation for speech reproduction

Peifeng Ji[1]*, Ee-Leng Tan[1], Woon-Seng Gan[1], *Senior Member, IEEE*, and Jun Yang[2], *Senior Member, IEEE*

[1] Digital Signal Processing Lab, Nanyang Technological University, Singapore 639798

[2] The State Key Laboratory of Acoustics and Key Laboratory of Noise and Vibration Research Institute of Acoustics, Chinese Academy of Sciences, Beijing, China 100190

*Corresponding Author Email: pfji@ntu.edu.sg

Abstract — Based on the Berktay's farfield solution, various preprocessing methods were proposed to reduce the distortion of the highly directional audible signal in the parametric loudspeaker. However, the Berktay's farfield solution is an approximated model of nonlinear acoustic propagation. To determine the effectiveness of these methods, we analyze various preprocessing methods theoretically for directional speech reproduction using the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation, which provides a more accurate model of nonlinear acoustic propagation. In order to reduce the distortion effectively in the parametric loudspeaker with these preprocessing methods, the initial sound pressure level of the carrier frequency is found to be less than 132 dB according to the KZK equation. Unlike the Berktay' farfield solution that results in a +12 dB/octave gain slope, different gain slopes are derived using the KZK equation and appropriate equalizers are proposed to improve the frequency response of the parametric loudspeaker. The optimal preprocessing method for directional speech reproduction is established based on the KZK equation, which has a relatively flat frequency response of the desired speech signal and the best total harmonic distortion performance.

Index Terms- Berktay's farfield solution, Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation, parametric array，speech reproduction, total harmonic distortion

EDICS: SPE-SPRD

I. INTRODUCTION

Directivity of the conventional loudspeaker is determined by its aperture and the frequency of operation. To achieve high directivity in low frequencies, the aperture size of the conventional loudspeaker must be very large. For example, to project a 100 Hz sound beam, the aperture size would be of several meters. By utilizing the nonlinear behavior of the air, parametric loudspeaker can be used to create loudspeakers having higher directivity but at a significantly smaller aperture as compared to the conventional loudspeaker. The parametric loudspeaker is generally referred as the parametric array deployed in air [1-4], which has been used in underwater sonar applications for decades.

The original work of the parametric array by Westervelt [5] discussed the generation of difference frequency signal from two high-frequency collimated beams, which are referred as primary waves. The directivity of the parametric array is attributed to an end-fire array of virtual sources that are the byproducts of the nonlinear interaction of the primary waves in the medium. A more complete explanation of the parametric array was given by Berktay [6] and his analysis was not limited to two primary waves. He used the concept of the modulation envelope and assumed that the primary wave has the form of $p_1 = P_0 E(t) \sin \omega_c t$, and obtained a farfield solution of

$$p_2 \propto P_0^2 \partial^2 E^2(t) / \partial t^2 , \tag{1}$$

where t is the time, $E(t)$ is the envelope function, P_0 is the initial sound pressure level (SPL) of the carrier, ω_c is the angular carrier frequency and p_2 is the demodulated signal. The Berktay's farfield solution, which defines the demodulated secondary waveform generated along the axis of the beam, is proportional to the second-time derivative of the square of the envelope. Consequently, an inherent +12 dB/octave gain is found in the demodulated signal. However, equalizing the entire audio spectrum to achieve a relatively flat frequency response is impractical [7]. In this study, we only equalize the speech spectrum (0.5 kHz ~ 3.5 kHz) by introducing an inverse slope gain to the input signal.

There are several commonly used amplitude-modulation based preprocessing methods that are currently employed to generate the envelope $E(t)$ in the parametric loudspeaker. An early attempt by Yoneyama *et al.* [2] used a double side band amplitude modulation (DSB-AM) method to generate an envelope function $E(t) = 1 + mg(t)$, where $g(t)$ is the input signal and m is the modulation index. From their experiments, they realized that the second harmonic is of similar amplitude to the fundamental signal and led to a high total harmonic distortion (THD). Based on the Berktay's solution, many attempts to improve the quality of the demodulated signal have been carried out using different preprocessing methods [7-14]. In 1984, Kamakura *et al.* [8] proposed a square root AM (SRAM) method that applies an envelope function $E(t) = \sqrt{[1 + mg(t)]}$. This envelope function has been verified to produce lower THD values as compared to the conventional DSB-AM method. However, this method requires ultrasonic emitter with very large bandwidth to reproduce the infinite harmonics introduced by the square root operation. Subsequently, Kamakura *et al.* [9] used another method which is based on the single-sideband amplitude modulation (SSB-AM) to reduce the distortion of the parametric loudspeaker. The main advantage of the SSB-AM method is that it only requires half the bandwidth of the DSB-AM method and produces an envelope that is similar to the one in the SRAM method, without the requirement of a large bandwidth ultrasonic emitter. In the special case of two primary waves, the SSB-AM method produces the exact same envelope as the SRAM method. However, for multiple primary waves, this is not the case and the envelope error occurs. Hence, a recursive SSB-AM method was proposed by Croft *et al.* [7, 10] to approximate the same envelope as in the SRAM method. Due to the high complexity of its recursive nature, a high-speed processor must be used to achieve real-time performance for the recursive SSB-AM method. Tan *et al.* [14] proposed a class of hybrid AM and SRAM methods which employ an orthogonal amplitude modulation. This method is called the modified AM (MAM) method and has the flexibility of controlling the amount of THD with different complexity and bandwidth requirements. This class of MAM preprocessing method also possesses lower complexity as compared to the recursive SSB-AM method.

Several THD analyses based on the Berktay's solution have been carried out in [13, 14], but without examining the effect of equalization. However, it should be noted that the Berktay's solution is an approximated solution of the sound beam propagation for weak nonlinearity [6]. Merklinger [15] gave a more general solution for the parametric array as $p_2 \propto P_0 \partial^2 \left[E(t) \tan^{-1}\left(\beta\omega_c P_0 E(t) / 4\alpha\rho_0 c_0^3 \right) \right] / \partial t^2$, where c_0 is the small-signal sound speed, ρ_0, α, β are the ambient density, the absorption coefficient of the carrier frequency [16], and the nonlinearity coefficient of the medium, respectively. For the case of $P_0 \ll 4\alpha\rho_0 c_0^3 / \beta\omega_c$, the Merklinger's solution simplifies to $p_2 \propto P_0^2 \partial^2 E^2(t) / \partial t^2$ which is the same as the Berktay's solution. For the case of $P_0 \gg 4\alpha\rho_0 c_0^3 / \beta\omega_c$, the Merklinger's solution indicates that $p_2 \propto P_0 \partial^2 \left| E(t) \right| / \partial t^2$. Both the Berktay's and Merklinger's solutions ignore the sound absorption of the difference frequency wave and the higher-order effects [15, 17]. Therefore, a more accurate model is required to account for the differences in performance of the various preprocessing methods [2, 7-10, 14] in the parametric loudspeaker.

It is well known that the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation [18] can accurately describe the propagation of finite amplitude sound beams by combining the effects of diffraction, absorption, and nonlinearity under a parabolic approximation. The validity of the KZK equation for the parametric array has been verified by many researchers [17, 19], and the Berktay's solution can be regarded as an approximation of the KZK equation for weak nonlinearity [17]. Although an analytical solution using the Gaussian-beam expansion technique has been proposed in [20-22] in the case of weak nonlinearity, there is still no explicit analytical solution to the KZK equation. In view of this limitation, various numerical solutions [23-26] have been developed in the frequency-domain, time-domain and combined time-frequency domain. The advantages of using the time-domain KZK algorithm are (i) its ability to compute arbitrary time waveform, and (ii) computational efficiency [27] as compared to other approaches.

In this paper, the time-domain KZK algorithm [25] is adopted to evaluate the THD performance of various preprocessing methods, namely DSB-AM, SRAM, SSB-AM and MAM, with and without equalization. In the case of the THD analysis, the recursive SSB-AM method is equivalent to the SSB-AM method and these two methods have the same THD performance. Hence, only the SSB-AM method will be investigated in this paper. The effectiveness of these preprocessing methods to reduce the distortion in the parametric loudspeaker is highly related to P_0 according to the Merklinger's solution [15]. Therefore, in this study based on the KZK equation, the effects of P_0 on the THD and the demodulated signals are investigated to determine an effective region of P_0 for these preprocessing methods. Moreover, two settings of P_0s in the effective region are chosen to determine which preprocessing method is the most applicable and suitable for the directional speech reproduction.

The rest of this paper is organized as follows. In Section II, the time-domain algorithm of the KZK equation, the performance index of THD, and the preprocessing methods are briefly introduced. Based on the KZK equation, Section III outlines the effects of P_0 on the THD for these four preprocessing methods with and without equalization. Discussions on the simulation results obtained using the KZK equation and the Berktay's solution are also presented. Finally, summary and conclusions are presented in Section IV.

II. THEORY

A. Description of the KZK Equation and its Time-Domain Algorithm

Consider the KZK nonlinear parabolic wave equation [26]:

$$\frac{\partial p}{\partial z}=\frac{c_0}{2}\int_{-\infty}^{\tau}\left(\frac{\partial^2 p}{\partial r^2}+\frac{1}{r}\frac{\partial p}{\partial r}\right)d\tau'+\frac{\beta p}{\rho_0 c_0^3}\frac{\partial p}{\partial \tau}+\frac{\delta}{2c_0^3}\frac{\partial^2 p}{\partial \tau^2}+\sum_{v}\frac{c'_v}{c_0^2}\int_{-\infty}^{\tau}\frac{\partial^2 p}{\partial \tau'^2}e^{-(\tau-\tau')/t_v}d\tau', \qquad (2)$$

where p is the sound pressure, z is the coordinate along the axis of beams, $\tau = t - z/c_0$ is the retarded time and δ is dissipation factor corresponding to thermoviscous absorption. Each relaxation process v (where v = 1, 2, …) is characterized by a relaxation time t_v and a small-signal sound speed increment c'_v. The atmosphere can be modeled as a thermoviscous fluid with two relaxation processes. The four terms on the right-hand side of (2) account for diffraction, nonlinearity, thermoviscous absorption and relaxation effect, respectively. After applying a coordinate transformation to the KZK equation, each term of (2) is solved individually at sufficiently small interval in time-domain. A detailed explanation of the time-domain algorithm to the KZK equation is given in [25] and [26].

B. Performance Index

In order to measure the performance of different preprocessing methods, the performance index THD is used. For a single angular frequency ω_1, THD is defined as follows:

$$\mathrm{THD}=\sqrt{\frac{T_2^2+\cdots+T_i^2+\cdots+T_n^2}{T_1^2+T_2^2+\cdots+T_i^2+\cdots+T_n^2}}\times 100\%, \qquad (3)$$

where T_1 and T_i represent the fundamental angular frequency component at ω_1, and higher harmonic components for i = 2, 3, …, n at angular frequency $i\omega_1$, respectively.

C. Preprocessing Methods

The block diagrams of four preprocessing methods (DSB-AM, SRAM, SSB-AM and MAM) are shown in Figs. 1-4. Table I lists the demodulated signal of each preprocessing method derived from the Berktay's solution and its corresponding THD expression when a single tone input, $g(t) = \cos\omega_1 t$ is considered. Out of the four preprocessing methods listed in Table I, only the DSB-AM method generates a demodulated signal that consists of both the difference frequency component (desired signal) ω_1 and the second harmonic $2\omega_1$. The rest of preprocessing methods only generate the difference frequency component ω_1 in the demodulated signal. Hence, only the DSB-AM method produces a THD value that is related to the modulation index, while the rest of the preprocessing methods produces a perfect THD performance. Therefore, THD result derived from the Berktay's model is not very useful as we are unable to differentiate the performance of the preprocessing methods, except for the DSB-AM method. A more complex KZK model will be used in the next section to provide a more detailed analysis in this study.

In Table I, the expression of the THD for the DSB-AM method [13] predicted by the Berktay's solution is $m/\sqrt{m^2+1} \times 100\%$, which is only increases monotonically with modulation index and independent on the different frequency component. Since Berktay predicted that the demodulated signal is proportional to the second-time derivative of the envelope, a +12 dB/octave gain slope is found in all demodulated signals in Table I. To equalize this inherent high pass filtering effect (+12 dB/octave, marked as the dashed line in Fig. 5) in the difference frequency component predicted by the Berktay's model, an inverse -12 dB/octave filter (marked as the dotted line in Fig. 5) is introduced to the input signal. Here, the difference frequency component spans from 0.5 kHz to 3.5 kHz with an interval of 500 Hz. Therefore, the effect of equalization is equivalent to lowering the modulation index when the difference frequency component increases [2]. An additional modulation index of $m' = 1/f_N^2$ can be introduced to these preprocessing methods to take into account of equalization, where f_N is the normalized difference frequency component with reference to the lowest frequency (i.e., 500 Hz in our study). Ideally, a flat frequency response of the difference frequency component can be obtained (marked as solid line in Fig. 5). For example, the envelope function of the DSB-AM method is $E(t) = 1 + mm' g(t)$ and the THD of the equalized DSB-AM method is expressed as

$$\frac{1}{\sqrt{\left(f_N^2/m\right)^2+1}} \times 100\% \,. \tag{4}$$

After equalization, the THD values in (4) not only increases monotonically with the modulation index, but also decreases monotonically with the difference frequency component.

III. THD ANALYSIS BASED ON THE KZK EQUATION

In this section, the demodulated process in the parametric loudspeaker for speech reproduction is investigated and the performance of the four preprocessing methods is analyzed using the THD results obtained from the KZK equation. The following parameters are used in our simulations. The radius of the ultrasonic emitter is 0.1 m, the carrier frequency is 40 kHz, the temperature and the relative humidity are set at 28ºC and 60%, respectively. The observation point is on axis and located at 1.8 m from the source. Same difference frequency components within the speech bandwidth are used the KZK analysis. Six modulation indices of m = 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 are also used in the preprocessing methods to cover the operating range of the parametric loudspeaker.

Assume that the ultrasonic emitter has a relatively flat response from 24 kHz to 56 kHz and can be approximated with an ideal bandpass filter in our simulations. The input speech signal is preprocessed and subsequently bandlimited by the above bandpass filter. The filtered signal is then applied to the KZK equation to model the nonlinear propagation process in air.

The DSB-AM method is generally regarded as the conventional method used in the parametric loudspeaker. Based on the Berktay's solution, many other methods such as SRAM, SSB-AM and MAM were proposed. These Berktay-based methods attempt to reduce the distortion found in the DSB-AM method by compensating the effect due to squaring of the envelope predicted by the Berktay's solution, and assuming that the second-time derivation effect can be compensated by the frequency response of the ultrasonic emitter. In this study using the KZK equation, we examine the effectiveness of the Berktay-based methods (SRAM, SSB-AM and MAM) by benchmarking their THD values against those obtained from the DSB-AM method. The SSB-AM method includes either lower SSB or upper SSB, and only the upper SSB is used in our study. Due to the narrow bandwidth of speech signals, there is not much difference in terms of attenuation and saturation effects [7] for using either cases.

A. Effect of Initial Sound Pressure Level P_0

Using the parameters defined in our simulations, the constant $4\alpha\rho_0 c_0^3 / \beta\omega_c$ in the Merklinger's solution [15] is found to be around 130 dB [7]. Hence, to determine the valid region of the Berktay's solution, we shall investigate the effect of P_0 varying from 100 dB to 140 dB at an interval of 2 dB. Also, the difference frequency component and m are defined as 2 kHz and 0.9, respectively.

Figure 6 illustrates the relation between the SPL of the difference frequency component at 2 kHz and P_0 using the four preprocessing methods based on the KZK equation. It is observed that when P_0 is greater than 132 dB, the SPL of the difference frequency component becomes approximately proportional to P_0. In contrast, when P_0 is reduced below 132 dB, we observe that the SPL of the difference frequency component is almost proportional to P_0^2. This observation agrees well with the Merklinger's solution, which predicts a threshold of around 130 dB. We also observe from Fig. 6 that the highest SPL of the difference frequency component at 2 kHz is obtained using the DSB-AM method, follows by the SSB-AM, MAM, and SRAM methods. The reasons can be elaborated as follows. The reproduction of 2 kHz can be simply regarded as the summation of nonlinear interaction of every pair of primary waves with a difference frequency of 2 kHz. For a nonlinear interaction of a pair of primary waves f_1 and f_2, where $f_1 > f_2$, and the amplitude of the difference frequency component $p_{f_1-f_2}$ is proportional to the product of primary wave amplitudes p_1, p_2, i.e., $p_{f_1-f_2} \propto p_1 p_2$ [18]. Also, the sound absorption of the pair of primary waves satisfies the condition: $\alpha_{f_2} < \alpha_{f_1}$ [16]. Table II summarizes the normalized amplitudes of the two pairs of primary waves that have the highest contribution in reproducing the difference frequency component of 2 kHz. Since there is only one sideband in the SSB-AM method resulting in only one pair of primary waves, the amplitude of 42 kHz primary wave from the SSB-AM method is divided into two equal components to facilitate the comparison in Table II. Several observations can be made from Table II. First, the contribution of primary wave pairs of the SRAM and MAM methods are relatively smaller than those in the DSB-AM method. Thus, it is expected that the difference frequency component obtained from the DSB-AM method is relatively higher in terms of SPL as compared to those from the SRAM and MAM methods. Second, for the DSB-AM and SSB-AM methods, the first and second primary wave pairs are of the same amplitude. However, the second primary wave pair of the SSB-AM method is of higher frequency as compared to the DSB-AM method. Considering the sound absorption [16], this results in the DSB-AM method producing a difference frequency with higher amplitude than the SSB-AM method.

Figure 7 illustrates the THD performance versus P_0 for the four preprocessing methods derived from the KZK equation. From Fig. 7, we note that the THD values for the SRAM, SSB-AM and MAM methods are significantly lower than the DSB-AM method when P_0 is less than 132 dB. This trend is reversed when P_0 is higher than 132 dB. This is expected as the Berktay's solution does not hold for P_0 higher than 132 dB, where the effectiveness of the Berktay-based preprocessing methods (SRAM, SSB-AM and MAM) is drastically degraded. Therefore, in this paper, we shall only consider P_0 lesser than 132 dB. As mentioned earlier, the major distortion of the DSB-AM method comes from the second harmonic. Interestingly, when the sound absorption [16] is taken into account in our KZK simulations, the SPL of the second harmonic is found to be significantly smaller than the difference frequency component [2]. The additional absorption term in the KZK equation leads to lower THD values as compared to those observations obtained from the Berktay's solution. Figure 6 shows the relation between the SPL of the second harmonic (4 kHz) of the DSB-AM method (marked as solid line with circle) and P_0, which exhibits similar trend but with a relatively slower variation as compared to the difference frequency component of 2 kHz. Therefore, a gradual reduction of the THD values for the DSB-AM method is observed as P_0 increases, as shown in Fig. 7.

We further investigate the relation between the frequency response of the difference frequency component and P_0 for each preprocessing method. To achieve a relative large SPL of the difference frequency component with relatively low THD, P_0 is chosen to vary from 112 dB to 128 dB at an interval of 4 dB based on the above analysis. The normalized frequency response of the difference frequency component using the preprocessing methods (m = 0.9) for different P_0s and preprocessing methods are shown in Fig. 8. A better fit to the results based on the KZK equation falls between +6 and +12 dB/octave, which also coincides with experimental results reported in [28]. When P_0 is smaller, the slope gain is closer to +12 dB/octave, and the slope gain gradually shifts to +6 dB/octave for higher P_0. These trends are observed for all preprocessing methods, as shown in Figs. 8(a)-(d); and also observed for other modulation indices, which are not shown in these plots for simplicity. Hence, the previous +12 dB/octave slope gain predicted by the Berktay's solution is inadequate for describing the frequency response of the difference frequency component under all P_0 settings. Therefore, a P_0-dependent equalizer must be designed to achieve a relatively flat frequency response.

Two representative settings of P_0s: P_0 = 128 dB (near $4\alpha\rho_0 c_0^3/\beta\omega_c$) and P_0 = 120 dB (much smaller than $4\alpha\rho_0 c_0^3/\beta\omega_c$) are chosen to compare the frequency response of the difference frequency component for the same P_0, as shown in Fig. 9. For the same P_0, the frequency responses are almost the same for these preprocessing methods. At P_0 = 128 dB (Fig. 9(a)), the frequency responses derived from the KZK equation are closer to +8.4 dB/octave slope. At lower P_0 = 120 dB (Fig. 9(b)), the frequency responses approach a higher slope of +9.6 dB/octave. For simplicity, the normalized frequency response of the second harmonic of the DSB-AM method is only shown in the inset of Fig, 9(a) for P_0 = 128 dB. It is worth noting that similar frequency response between the second harmonic of the DSB-AM method and the difference frequency component using each preprocessing method is found. This finding is consistent with our observations in Section II.

B. THD Simulations based on the KZK Equation

Figure 10 shows four subplots that correspond to the THD performance obtained by the four preprocessing methods without equalization at P_0 = 128 dB. Six different modulation indices (from m = 0.5 to 1 at an interval of 0.1) are examined over the speech frequency range. In these simulations, a THD of 10% is used as a performance indicator to distinguish between poor and acceptable performance. A THD of 10% and larger is considered as poor performance, and vice versa. The THD values for these four preprocessing methods increase monotonically with the modulation index. Although they still exhibit lower THD values than the DSB-AM method, the THD values (derived from the KZK model) of these Berktay-based preprocessing methods are greater than 7.5%. It is attributed to the relatively high P_0 of 128 dB. For this level of P_0, the Berktay's solution is unable to accurately predict the propagation of the sound beam. Therefore, this inaccuracy causes all the Berktay-based preprocessing methods to exhibit higher harmonic distortion at high P_0. It is also difficult to predict the characteristics of the distortions as well as the THD values of these Berktay-based methods at such high P_0. But for the DSB-AM method, similar performance is still obtained from both the KZK and Berktay's models, i.e., the THD is almost independent of the difference frequency component, due to the fact that the significant distortion of the DSB-AM method comes from the second harmonic.

The THD performances of the preprocessing methods without equalization are summarized in Table III. The percentage in Table III defines the ratio of the number of the combined THD values smaller than 10% over the total number of the combined THD values across the speech frequency range and accumulated over the six modulation index settings. Therefore, a higher percentage implies that a smaller amount of distortion is obtained. From Table III, it is clear that the best preprocessing method without equalization is the SRAM method. However, the percentages of the THD values lower than 10% in the SRAM, SSB-AM and MAM methods without equalization are relatively small.

So we can conclude that these three Berktay-based methods are unable to effectively reduce harmonic distortion for P_0 = 128 dB without equalization.

For simplicity and clarity, in the remaining of this paper, only the THD values for m = 0.9 will be plotted to illustrate the performance of these four preprocessing methods. But to compare the overall performance of these preprocessing methods, the THD values for all six modulation indices are still included to calculate the percentage of the THD values lower than 10%.

The uneven frequency response of the difference frequency component derived from the KZK model, which uses a P_0 = 128 dB with a +8.4 dB/octave slope gain in Fig. 9(a), is undesirable for speech reproduction. Therefore, an equalizing filter with a slope of –8.4 dB/octave should be applied to the input speech signal, which is similar to equalization shown in the Fig. 5. After equalization is applied, the normalized frequency response of the difference frequency component using the preprocessing methods is shown in Fig. 11 for m = 0.9. From Fig. 11, the frequency response of the difference frequency component with equalization becomes relatively flat and exhibits a ripple of less than 1.5 dB. However, it should be noted that equalization of the input signal lowers the electric-acoustic conversion efficiency. For the DSB-AM method, the SPL of the second harmonic is proportional to m^2 and the frequency response of the second harmonic is approximated as a slope of +8.4 dB/octave without equalization, as shown in the inset of Fig. 9(a). As mentioned earlier that the effect of equalization is equivalent to reducing the modulation index with respect to the difference frequency component [2]. Hence, the second harmonic of the DSB-AM method is expected to be attenuated by around (–2×8.4 + 8.4) dB/octave with equalization. This finding has been verified with the normalized frequency response of the second harmonic of the DSB-AM method as shown in the inset of Fig. 11.

Using the KZK equation, the THD values of the four preprocessing methods with equalization are illustrated in Fig. 12 for m = 0.9. Equalizing the speech signal has the effect of lowering the modulation index with respect to the difference frequency component [2], therefore, the THD values with equalization decrease monotonically with the difference frequency component. This observation derived from the KZK equation is also similar to that from (4). The THD performances of the preprocessing methods with equalization are also summarized in Table III. The percentages for all preprocessing methods increase significantly after equalization, which implies that equalization is crucial for all preprocessing methods operating at P_0 = 128 dB. It is also interesting to note that the SRAM and MAM methods provide the best THD performance with equalization. The MAM method improves its THD performance significantly after equalization. Therefore, for a relatively high P_0, it is highly recommended to adopt equalization to these Berktay-based methods.

Similar to the previous case, we now examine the case of lower P_0 of 120 dB. Figure 13 shows the THD values obtained using the four preprocessing methods without equalization for m = 0.9. For the DSB-AM method, the distortion is almost independent of the difference frequency component due to the same reason mentioned for the case of P_0 = 128 dB. From Fig. 13, it is noted that the THD values for the SRAM, SSB-AM and MAM methods are smaller as compared to the DSB-AM method for all frequencies. The poor frequency response of the parametric loudspeaker at low frequencies causes poorer THD performance, where a relatively higher noise floor is generated and results in a lower signal-to-noise ratio (SNR). The noise floor comes from the nonlinear acoustic propagation modeling of finite amplitude wave and also due to the limited time window length adopted in the FFT analysis. The THD values of the SRAM, SSB-AM and MAM methods are approximately proportional to the difference frequency component as the SPL of the difference frequency component increases by a factor between +6 and +12 dB/octave [28], which is much larger than the increment of its harmonics. In the case of lower P_0, the THD values of the SRAM, SSB-AM and MAM methods indicate that the Berktay's solution can well predict the propagation of the sound beam; hence, the Berktay-based preprocessing methods can be employed to reduce distortion effectively.

Using the KZK model, the THD values of the preprocessing methods with equalization are obtained and shown in Fig. 14 for $m = 0.9$. The equalizing filter used in these methods has a slope of –9.6 dB/octave. The THD values for the DSB-AM method with equalization decrease monotonically with the difference frequency component, which is consistent with the trend predicted from (4). There are some fluctuations at higher difference frequency components. These fluctuations are caused by the noise floor as it is relatively larger as compared to the second harmonic at such conditions and becomes the highest contribution to distortion. For the Berktay-based methods, most of the THD values become higher due to the lower SNR after equalization is applied. This is attributed to the higher reduction of the difference frequency component as compared to the noise floor and the harmonics. Therefore, equalization is not necessary for lower P_0.

With equalization and setting $m = 0.9$, the normalized frequency response of the difference frequency component using the preprocessing methods is relatively flat with ripple less than 1.6 dB, as shown in Fig. 15. A summary of the THD performance of the preprocessing methods for all six modulation indices is presented in Table IV. From Table IV, it is clear that the best methods without equalization are the SRAM and SSB-AM methods, and the best method with equalization is the SSB-AM method.

IV. SUMMARY AND CONCLUSIONS

Using the KZK equation, a comparative theoretical analysis for four preprocessing methods, which are commonly used in the parametric loudspeaker to reproduce the directional speech signal, has been investigated with and without equalization. In this paper, we focus on comparing the simulation results obtained from the demodulated signals between using the KZK equation and the Berktay's solution. After determining the relation between the THD and P_0 according to the KZK equation, P_0 is limited to less than 132 dB in this study. The gain slopes of the frequency response of the difference frequency component obtained from the KZK equation are smaller than that predicted from the Berktay's solution and are between +6 and +12 dB/octave. Based on the gain slopes from the KZK equation, the equalizing filters for the preprocessing methods are designed and subsequently applied to the input speech signal for the cases of P_0 = 128 dB and 120 dB, and the equalized frequency responses of the difference frequency component are relatively flat and exhibit a ripple of less than 1.5 dB and 1.6 dB, respectively.

In the case of P_0 = 128 dB, the nonlinear effect is relatively stronger, but the Berktay-based preprocessing methods (SRAM, SSB-AM and MAM) still exhibit slightly lower THD values than the conventional DSB-AM method. There is also a drop of the THD values after equalization due to the higher SNR and the equivalent effect of equalization is to lower the modulation index with respect to the difference frequency component. In the case of P_0 = 120 dB, the Berktay-based preprocessing methods effectively reduce their THD values when compared to the conventional DSB-AM method. However, their THD values increase when the equalizing filter is applied primarily due to the lower SNR after equalization. By using a THD of 10% as a performance indicator, the best preprocessing methods are SRAM and SSB-AM for P_0 = 128 dB and P_0 = 120 dB, respectively.

From our analysis, the most applicable and suitable preprocessing methods to reproduce speech are SRAM and MAM for the case of P_0 = 128 dB with equalization. These methods lead to a relatively flat frequency response of the difference frequency component. Also, these methods have the best THD performance at the cost of lower electric-acoustic conversion efficiency. Although only the frequency band of speech signal has been analyzed theoretically in this study, similar analysis can be carried out at lower or higher frequencies. The experimental performance of these preprocessing methods when applied to the parametric loudspeaker will be reported in our next paper.

ACKNOWLEDGMENTS

This work is supported by the Singapore National Research Foundation Interactive Digital Media R&D Program, under research grant **NRF2007IDM-IDM002-086**.

REFERENCES

[1] M. B. Bennett and D. T. Blackstock, "Parametric array in air," *J. Acoust. Soc. Am.,* vol. 57, pp. 562-568, 1975.

[2] 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.*, vol. 73, pp. 1532-1536, 1983.

[3] W. S. Gan, J. Yang, K. S. Tan, and M. H. Er, "A digital beamsteerer for difference frequency in parametric array," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, pp. 1018-1025, 2006.

[4] K. Lee and W. S. Gan, "Bandwidth-efficient recursive *p*th-order equalization for correcting baseband distortion in parametric loudspeakers," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, pp. 706-710, 2006.

[5] P. J. Westervelt, "Parametric acoustic array," *J. Acoust. Soc. Am.*, vol. 35, pp. 535-537, 1963.

[6] H. O. Berktay, "Possible exploitation of non-linear acoustics in underwater transmitting

applications," *J. Sound Vib.*, vol. 2, pp. 435-461, 1965.

[7] J. J. Croft and J. O. Norris, "Theory, history, and the advancement of parametric loudspeaker," *White paper of American Technology Corporation*, 2001.

[8] T. Kamakura, M. Yoneyama, and K. Ikegaya, "Development of a parametric loudspeaker for practical use," in *10th International Symposium on Nonlinear Acoustics*, Kobe, 1984, pp. 147-150.

[9] T. Kamakura, Y. Tasahide, and K. Ikeyaya, "A study for the realization of a parametric loudspeaker," *J. Acoust. Soc. Japan*, pp. 1-18, 1985.

[10] J. J. Croft, M. E. Spencer, and J. O. Norris, "Modulator processing for a parametric speaker system," U. S. 20080063214A1, 2003.

[11] T. D. Kite, J. T. Post, and M. F. Hamilton, "Parametric array in air: distortion reduction by preprocessing," in *Proc.16th ICA and 135th Meeting of the ASA*, 1998, pp. 1091-1092.

[12] F. J. Pompei, "The use of airborne ultrasonics for generating audible sound beams," *J. Audio Eng. Soc.*, vol. 47, pp. 726-731, 1999.

[13]Y. H. Liew, "Signal processing techniques for sound reproduction in parametric array." Singapore: Nanyang Technological University, 2002.

[14] E. L. Tan, W. S. Gan, P. Ji, and J. Yang, "Distortion analysis and reduction for the parametric array," in *124th Convention of the Audio Engineering Society*, 2008.

[15]H. M. Merklinger, "Improved efficiency in the parametric transmitting array," *J. Acoust. Soc. Am.*, vol. 58, pp. 784-787, 1975.

[16] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Atmospheric absorption of sound: Further developments," *J. Acoust. Soc. Am.*, vol. 97, pp. 680-683, 1995.

[17] M. A. Averkiou, Y.-S. Lee, and M. F. Hamilton, "Self-demodulation of amplitude- and frequency-modulated pulses in a thermoviscous fluid," *J. Acoust. Soc. Am.*, vol. 94, pp. 2876-2883, 1993.

[18] M. F. Hamilton and D. T. Blackstock, *Nonlinear acoustics*. San Diego: Academic Press, 1998.

[19] T. Kamakura, N. Hamada, K. Aoki, and Y. Kumamoto, "Nonlinearly generated spectral components in the nearfield of a directive sound source," *J. Acoust. Soc. Am.*, vol. 85, pp. 2331-2337, 1989.

[20] J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," *J. Acoust. Soc. Am.*, vol. 83, pp. 1752-1756, 1988.

[21] J. Yang, K. Sha, W. S. Gan, and J. Tian, "Modeling of finite-amplitude sound beams: Second order fields generated by a parametric loudspeaker," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 52, pp. 610-618, 2005.

[22] H. J. Kim, L. W. Schmerr, and A. Sedov, "Generation of the basis sets for multi-Gaussian ultrasonic beam models - An overview," *J. Acoust. Soc. Am.*, vol. 119, pp. 1971-1978, Apr 2006.

[23] N. S. Bakhvalov, Y. M. Zhileikin, E. A. Zabolotskaya, and R. V. Khokhlov, "Nonlinear propagation of a sound beam in a nondissipative medium," *Sov. Phys. Acoust.*, vol. 22, pp. 272-274, 1976.

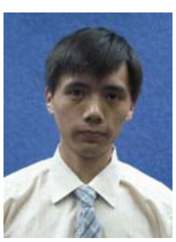
[24] S. I. Aanonsen, T. Barkve, J. N. Tjøtta, and S. Tjøtta, "Distortion and harmonic generation in the nearfield of a finite amplitude sound beam," *J. Acoust. Soc. Am.*, vol. 75, pp. 749-768, 1984.

[25] Y.-S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," *J. Acoust. Soc. Am.*, vol. 97, pp. 906-917, 1995.

[26] R. O. Cleveland, M. F. Hamilton, and D. T. Blackstock, "Time-domain modeling of finite-amplitude sound in relaxing fluids," *J. Acoust. Soc. Am.*, vol. 99, pp. 3312-3318, 1996.

[27] W. Kim and V. W. Sparrow, "Audio application of the parametric array implementation through a numerical model," in *113th Convention of the Audio Engineering Society*, 2002.

[28] I. O. Wygant, M. Kupnik, J. C. Windsor, W. M. Wright, M. S. Wochner, G. G. Yaralioglu, M. F. Hamilton, and B. T. Khuri-Yakub, "50 kHz capacitive micromachined ultrasonic transducers for generation of highly directional sound with parametric arrays," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, pp. 193-203, 2009.



Peifeng Ji received the B. Eng. and M. Eng. degrees from Shandong University of Science and Technology, Qingdao, China, and the Ph.D. degree from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, in 2002, 2005, 2008, respectively. Now he is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, as a research fellow. His research interests are mainly in acoustic signal processing, acoustical measurement and nonlinear acoustics.

Ee-Leng Tan received his B.Eng. (1st Class Hons) in electrical and electronics engineering from Nanyang Technological University (NTU), Singapore, in 2003. Currently, he is with NTU as a research associate. His research interests include image processing, nonlinear acoustics and real-time digital signal processing.

Woon-Seng Gan (S’90-M’93-SM’00) received his BEng (First Class Hons) and Ph.D. degrees, both in Electrical and Electronic Engineering from the University of Strathclyde, UK in 1989 and 1993 respectively. He joined the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, as a Lecturer and Senior Lecturer in 1993 and 1998 respectively. In 1999, he was promoted to Associate Professor. He is currently the Deputy Director of the Center for Signal Processing at Nanyang Technological University. His research interests include adaptive signal processing, psycho-acoustical signal processing, audio processing, and real-time embedded systems.

Dr. Gan has also won the Institute of Engineer Singapore (IES) Prestigious Engineering Achievement Award in 2001 for the work on Audio Beam System. He has published more than 160 international refereed journals and conferences, and has been awarded four Singapore and US patents. He has co-authored a book on *Digital Signal Processors: Architectures, Implementations, and Applications* (Prentice Hall, 2005). This book has since been translated to chinese for adoption by universities in China. He is also the leading author of a new book on *Embedded Signal Processing with the Micro Signal Architecture*, (Wiley-IEEE, 2007). A new book on *Subband Adaptive Filtering: Theory and Implementation* is due to be published by Wiley in July 2009.

Jun Yang (M’99-SM’04) received the B.Eng. and M. Eng. degrees from the Harbin Engineering University, Harbin, China, and the Ph.D. degree in acoustics from Nanjing University, Nanjing, China in 1990, 1993 and 1996, respectively. From 1996 to 1998, he was a postdoctoral fellow at Institute of Acoustics, Chinese Academy of Sciences (IACAS), Beijing, China. From October 1998 to April 1999, he worked in Hong Kong Polytechnic University as a visiting scholar. From Jan 1997 to May 1999, he was with IACAS as an associate professor. He joined the School of Electrical and Electronic Engineering (EEE), Nanyang Technological University (NTU), Singapore, as a research fellow, a teaching fellow, an assistant professor and an associate professor in 1999, 2001, 2003 and 2005, respectively. Since Nov 2003, he has been a professor at IACAS, China. Currently, he is the Director of the Communication Acoustics Laboratory at IACAS and the Deputy Director of the Key Laboratory of Noise and Vibration Research Institute of Acoustics, Chinese Academy of Science. His research interests include communication acoustics, 3-D audio system, acoustic signal processing, sound field control and nonlinear acoustics.

Table I

Demodulated signal and its THD for each preprocessing method based on the Berktay's solution (for single tone input).

Preprocessing method	Demodulated signal	THD
DSB-AM	$p_2 \propto \left[m\omega_1^2 \cos(\omega_1 t) + m^2\omega_1^2 \cos(2\omega_1 t) \right]$	$m / \sqrt{m^2+1} \times 100\%$
SRAM	$p_2 \propto \left[m\omega_1^2 \cos(\omega_1 t) \right]$	0
SSB-AM	$p_2 \propto \left[m\omega_1^2 \cos(\omega_1 t) \right]$	0
MAM	$p_2 \propto \left[m\omega_1^2 \cos(\omega_1 t) \right]$	0

Table II

Normalized amplitudes of the primary wave pairs to reproduce 2 kHz for each preprocessing method

Preprocessing method	Normalized amplitudes of 1st pair of primary waves		Normalized amplitudes of 2nd pair of primary waves	
	38 kHz	40 kHz	40 kHz	42 kHz
DSB-AM	0.45	1	1	0.45
SRAM	0.2696	1	1	0.2696
SSB-AM	0.45 (42 kHz)	1	1	0.45
MAM	0.3603	1	1	0.3603

Table III

Percentage of combined THD values lower than 10% at P_0 = 128 dB accumulated over m = 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0

Preprocessing method	Without equalization (%)	With equalization (%)
DSB-AM	0	42.86
SRAM	21.43	85.71
SSB-AM	4.76	71.43
MAM	7.14	85.71

Table IV

Percentage of combined THD values lower than 10% at P_0 = 120 dB accumulated over m = 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0

Preprocessing method	Without equalization (%)	With equalization (%)
DSB-AM	0	30.95
SRAM	85.71	26.19
SSB-AM	85.71	42.86
MAM	78.57	21.43

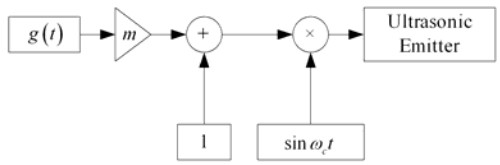


Fig. 1. Block diagram of the DSB-AM method.

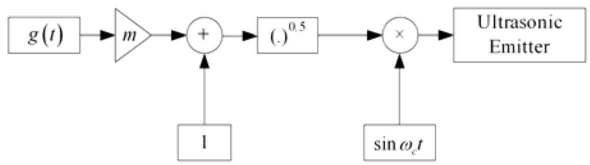


Fig. 2. Block diagram of the SRAM method.

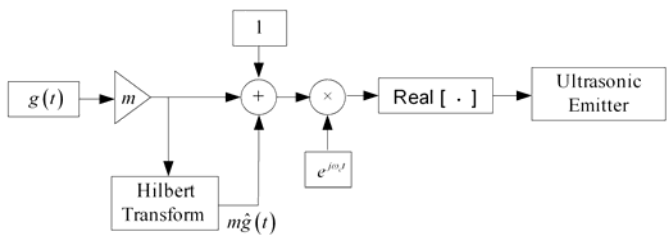


Fig. 3. Block diagram of the SSB-AM method, where $j = \sqrt{-1}$.

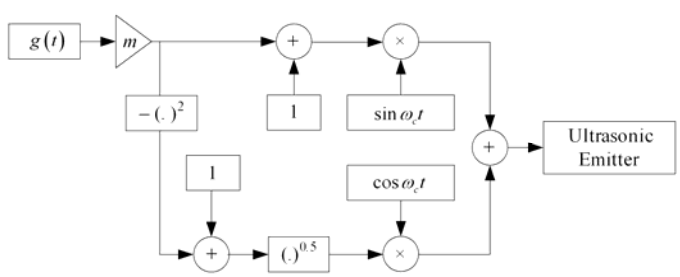


Fig. 4. Block diagram of the MAM method.

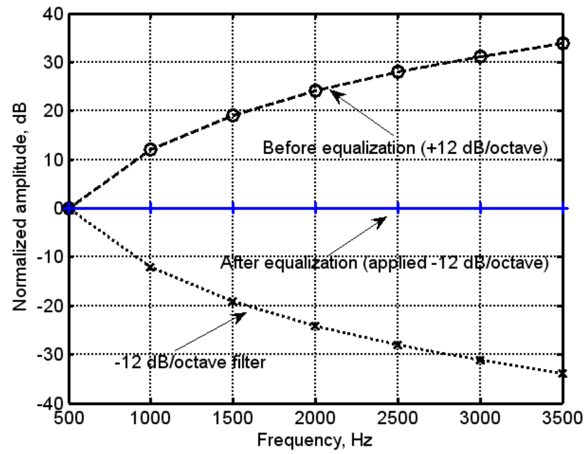


Fig. 5. Equalization of the difference frequency component.

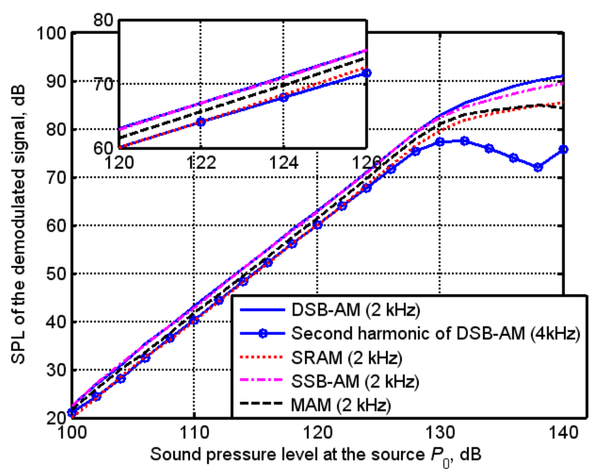


Fig. 6. Relation between SPL of the difference frequency component of each preprocessing method, second harmonic of the DSB-AM method and P_0 based on the KZK equation.

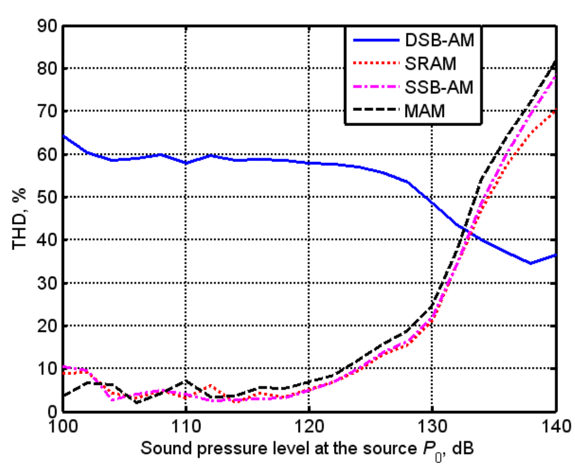


Fig. 7. Relation between THD and P_0 based on the KZK equation.

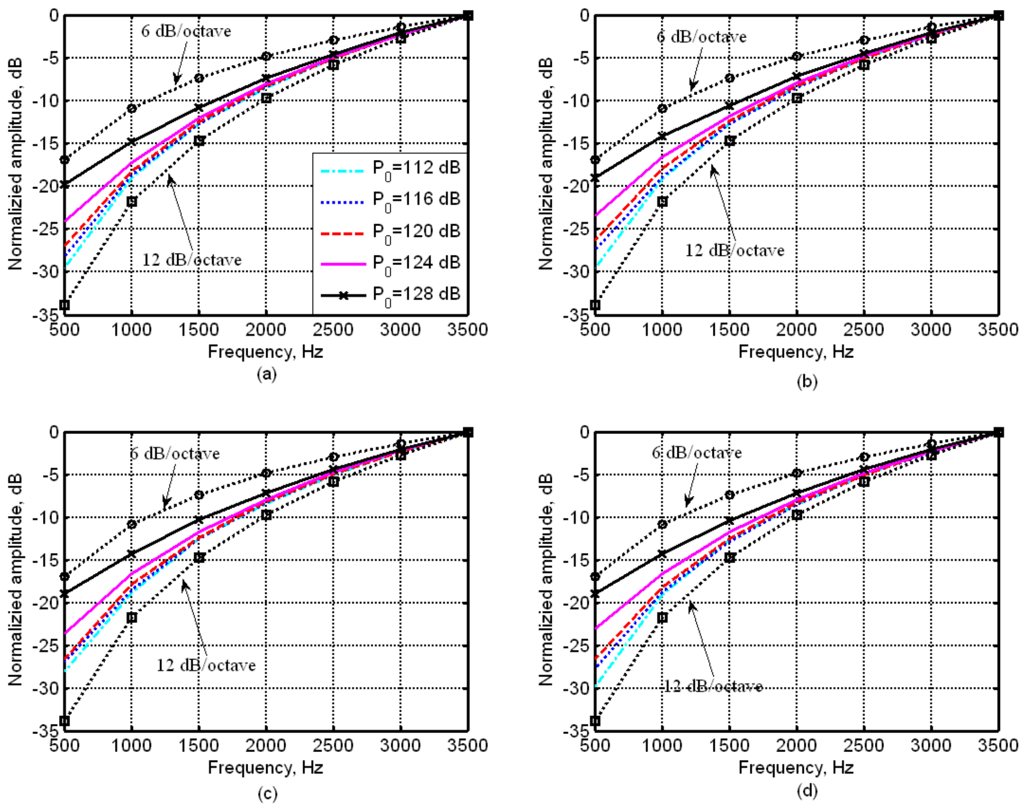
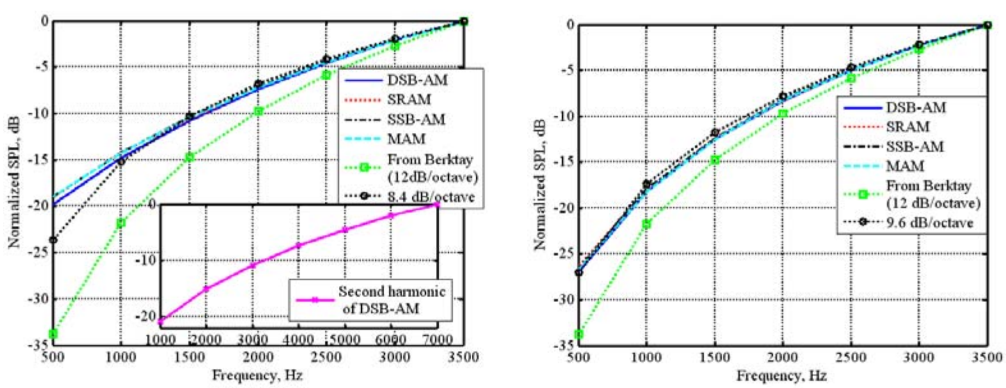


Fig. 8. Frequency response of the difference frequency component for different P_0s (m = 0.9) for (a) DSB-AM, (b) SRAM, (c) SSB-AM and (d) MAM. Same legend in (a) applies to all plots.



(a) (b)

Fig. 9. Normalized frequency response of the difference frequency component for four preprocessing methods when m = 0.9 for (a) P_0 = 128 dB and (b) P_0 = 120 dB.

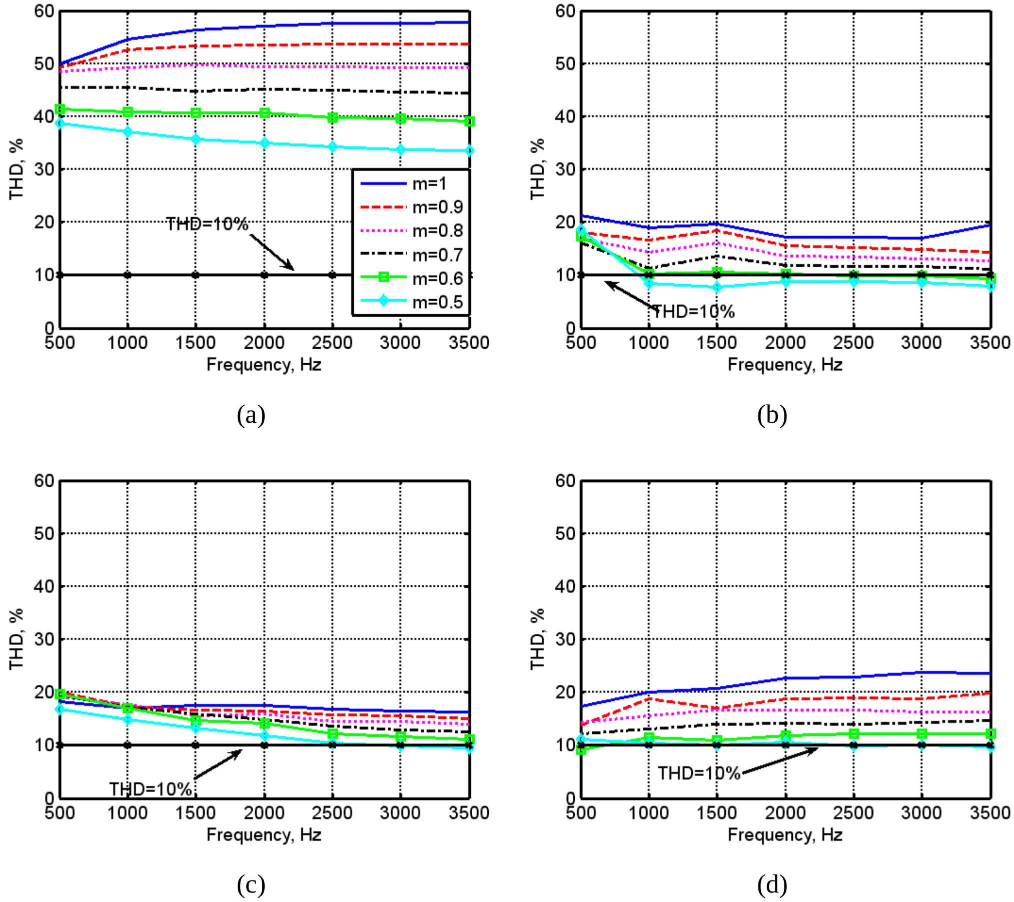


Fig. 10. THD values for preprocessing methods without equalization for P_0 = 128 dB with (a) DSB-AM, (b) SRAM, (c) SSB-AM and (d) MAM. Same legend in (a) applies to all plots.

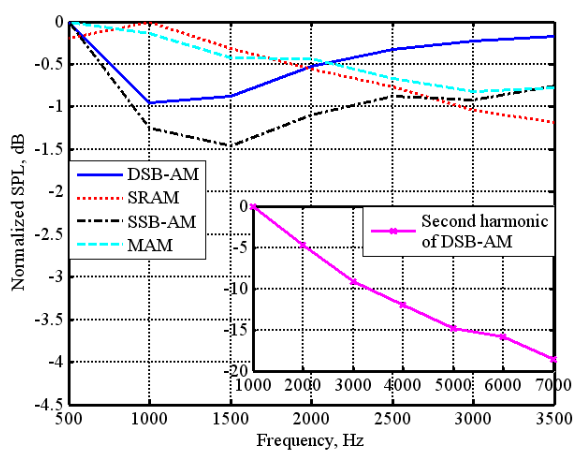


Fig. 11. Normalized frequency response of the difference frequency component for four preprocessing methods and second harmonic of DSB-AM method with equalization for P_0 = 128 dB and m = 0.9.

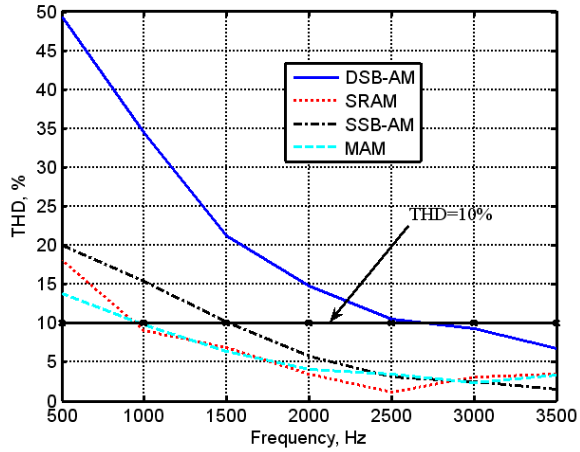


Fig. 12. THD values for preprocessing methods with equalization for P_0 = 128 dB and m = 0.9.

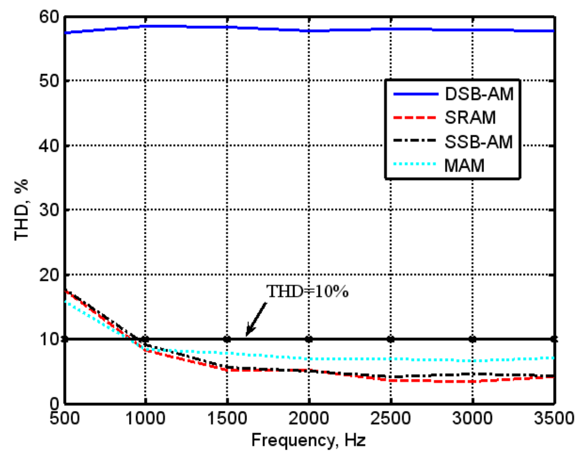


Fig.13. THD values for preprocessing methods without equalization for P_0 = 120 dB and m = 0.9.

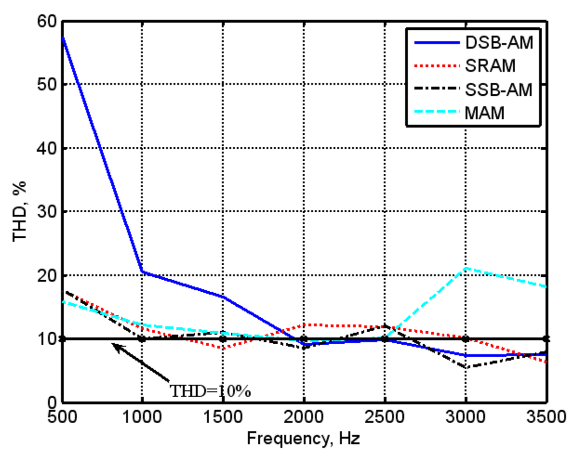


Fig.14. THD values for preprocessing methods with equalization for P_0 = 120 dB and m = 0.9.

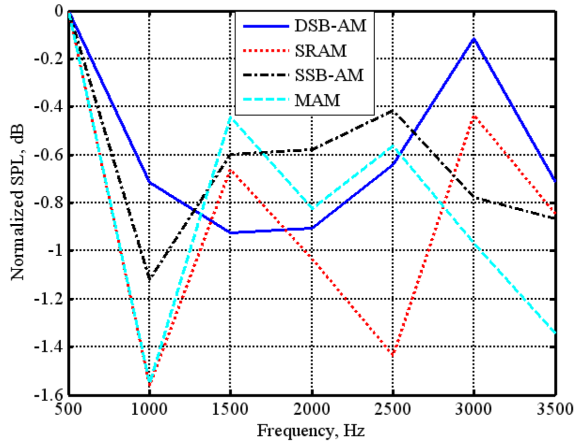


Fig.15. Normalized frequency response of the difference frequency component for four preprocessing method with equalization for P_0 = 120 dB and m = 0.9.