INTER-NOISE 2021 Congress
(1–4 August 2021, Online and in Washington DC, USA)

Virtual bass preprocessing and carrier frequency optimization for the parametric array loudspeaker

Shengqi Tao[1]
University of Electronic Science and Technology of China
2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu, China 611731

Jing Ren[2]
University of Electronic Science and Technology of China
2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu, China 611731

Chuang Shi[3]
University of Electronic Science and Technology of China
2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu, China 611731

ABSTRACT

The parametric array loudspeaker (PAL) is a novel type of loudspeaker that can project a directional sound beam. It is usually used in creating personal sound zone and projecting private messages to a targeted audience. However, the PAL has a very poor low-frequency response due to the inherent nonlinear acoustic principle generating sound from ultrasound in air. A psychoacoustic signal processing method known as the virtual bass (VB) has been proved to be an effective method to improve the bass quality of consumer electronics with miniature or flat loudspeaker unit. This paper proposes the VB preprocessing based on the phase vocoder (PV) for the bass enhancement of the PAL that adopts a vestigial sideband (VSB) modulation method. The harmonics generated by the VB preprocessing are presented in the lower sideband, while the high-passed audio input is embedded in the upper sideband. A measure, namely the in-band peak flatness, is thereafter proposed in this paper to determine the optimal carrier frequency, given a practical uneven frequency response of the ultrasonic emitter. The subjective test results validate that the proposed VB preprocessing together with the optimal carrier frequency can improve the bass and overall sound quality of the PAL.

1. INTRODUCTION

The PAL uses the parametric array effect in air to generate a hyper-directional sound beam [1]. When two primary waves, for example at 40 kHz and 42 kHz, propagate on the same path, their difference frequency wave at 2 kHz is gradually generated due to the nonlinearity of air [2]. Resulting

[1]taosq@std.uestc.edu.cn

[2]renj@std.uestc.edu.cn

[3]shichuang@uestc.edu.cn (corresponding author)

from the interaction between the two primary waves, the difference frequency wave exhibits a sharp directivity that is similar to those of the primary waves [3]. Thus, an ultrasonic carrier wave beyond human hearing range is adopted in the PAL to deliver the audible sound to a targeted location. The ultrasonic carrier frequency is usually set at the resonant frequency of the ultrasonic emitter. The audio input is modulated on the ultrasonic carrier wave and the parametric array effect can demodulate it to recover the audible frequencies. This self-demodulation process is readily explained by the aforementioned example when the two primary waves have very close frequencies and their difference frequency is presented by the envelope of the superposition of the two primary waves.

The PAL can be applied in many attractive applications, such as private messaging, spatial audio, active noise control, interactive sound visualization and so on [4, 5]. However, there are two major limitations of the PAL [6]. Firstly, in the self-demodulation process, harmonic distortion of the desired sound is generated as a by-product. Secondly, the self-demodulation process shows a high-pass filtering effect, resulting in a very poor low-frequency response of the PAL. Research works over the last two decades have mainly been focused on reducing the harmonic distortion and controlling the beam pattern of the PAL [7]. There have been few studies on solving the bandwidth limitation of the PAL [8]. In order to boost up the low-frequency response of the PAL, the total output power has to be increased significantly, which results in less conversion efficiency because of the saturation in air. Alternatively, a psychoacoustic phenomenon, namely the missing fundamental effect, has previously been investigated for the PAL to create the VB [9]. Human auditory system can perceive the fundamental frequency from its higher harmonics [10]. Therefore, if the PAL is able to transmit appropriate harmonics, its bass reproduction can be perceptually enhanced.

When the VB preprocessing is added in the PAL, the audio input is firstly split into two frequency bands through a pair of low-pass and high-pass filters. The lower band is processed by a harmonic generator to create the harmonics requested by the VB. At present, there are two choices to implement this harmonic generator. They are the nonlinear device (NLD) and the phase vocoder (PV) [11–13]. The NLD is simple but prone to nonlinear distortion, although the NLD has previously been validated to be effective in the PAL [8]. The PV outperforms the NLD in precise control over the number and amplitudes of the generated harmonics. This is essential for the PAL, as the PAL simultaneously generates harmonics from the parametric array effect.

2. THEORY AND METHOD

Figure 1 shows the proposed PV-based virtual bass preprocessing for the PAL. The harmonics generated by the PV is modulated on the lower sideband, while the high-passed audio input is modulated on the upper sideband. Together, they form a VSB modulation scheme. The signals of these two channels are mixed up as the input of the ultrasonic emitter.

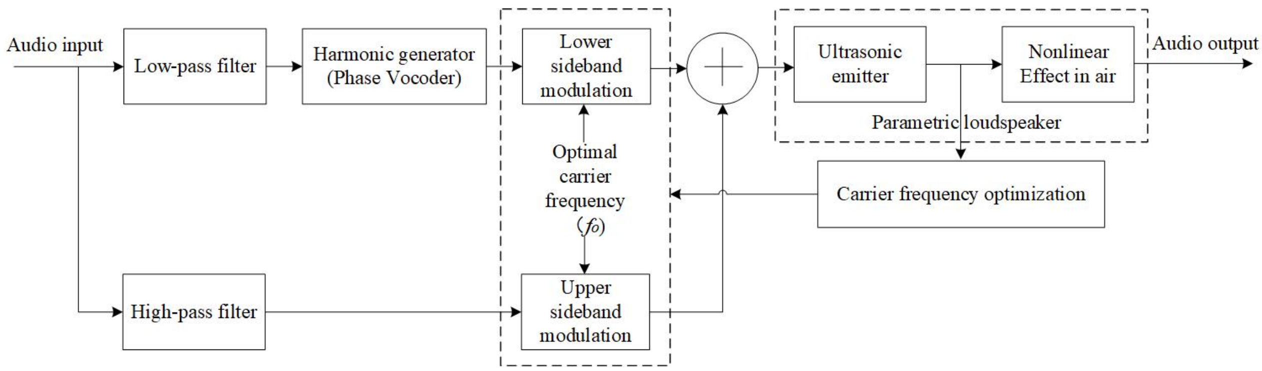

Figure 1: Block diagram of the virtual bass preprocessing and carrier frequency optimization for the PAL.

2.1. PV-based Virtual Bass Preprocessing

The PV uses pitch-shifting functions to generate harmonics [13]. The fundamental assumption of the PV is that the input signal can be modeled as a sum of slowly varying sinusoids, *i.e.*

$$x_{LP}(n) = \sum_{k=1}^{I(n)} A_k(n)\cos[\phi_k(n)] \tag{1}$$

and

$$\phi_k(n) = \phi_k(0) + \sum_{\tau=0}^{n} \frac{2\pi f_k(\tau)}{f_s}, \tag{2}$$

where $x_{LP}(n)$ denotes the low-passed audio input ; $\phi_k(n)$, $A_k(n)$ and $f_k(n)$ are the instantaneous phase, amplitude and frequency of the k-th sinusoid, respectively; $I(n)$ is the number of sinusoids; and f_s is the sampling frequency.

By using the short-time Fourier transform (STFT), the spectrogram of the input signal is obtained as

$$X_{LP}(m,k) = \sum_{n=0}^{L_{win}-1} x_{LP}(mR_a + n)\cdot h(n)\cdot exp(-j\frac{2\pi k}{L_{win}}n), \tag{3}$$

where m and k are the frame index and the frequency bin index, respectively; L_{win} and R_a are the frame width and hop size, respectively; and the window function is denoted as $h(n)$.

The spectrogram consists of the magnitude $|X_{LP}(m,k)|$ and the phase $\angle X_{LP}(m,k)$. In the PV, the amplitude of the pitch-shifted signal is determined by $|X_{LP}(m,k)|$. For example, the exponential weighting scheme is carried out by

$$|X_{PV}(m,k)| = W_i \cdot |X_{LP}(m,k)|, \tag{4}$$

where $W_i = exp(-0.3\cdot i)$. The instantaneous frequencies of the input signal are estimated based on $\angle X_{LP}(m,k)$. The phase of the pitch-shifted signal is determined by

$$\angle X_{PV}(m,k) = \angle X_{PV}(m-1,k) + \alpha\cdot 2\pi f_k(mR_a + L_{win}/2)\cdot\frac{R_a}{f_s}, \tag{5}$$

where α is the order of the harmonic. In the PAL, frequency components below 400 Hz are fed into the PV, and only their second and third harmonics are generated. Lastly, the harmonics are obtained from the inverse Fourier transform of the pitch-shifting signal.

2.2. Carrier Frequency Optimization

The single sideband (SSB) modulation is essentially a quadrature modulation method. It introduces a quadrature path that cancels nonlinear distortion occurred in the double sideband (DSB) modulation [14]. The quadrature path in the SSB modulation is usually implemented with the Hilbert transform, which is known as a broadband phase-shifted network.

This paper adopts a vestigial sideband modulation (VSB) method. The VSB modulation method consists of two SSB modulations. The harmonics generated by the PV are presented in the lower sideband, while the high-passed audio input is embedded in the upper sideband. The modulated signals are written as

$$y_{USB}(t) = \frac{1}{2}[x_{HP}(t)\cos(\omega_c t) - \hat{x}_{HP}(t)\sin(\omega_c t)] \tag{6}$$

and

$$y_{LSB}(t) = \frac{1}{2}[x_{PV}(t)\cos(\omega_c t) + \hat{x}_{PV}(t)\sin(\omega_c(t))], \tag{7}$$

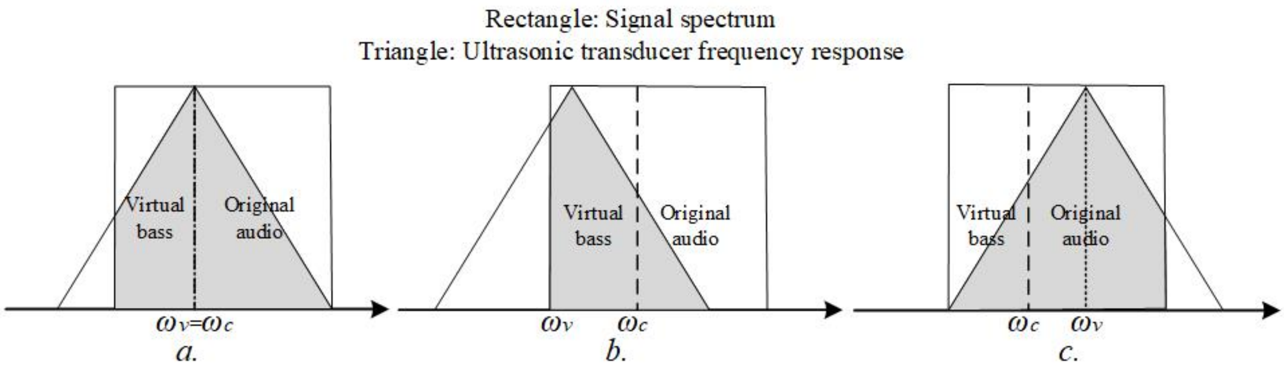

Figure 2: Illustration of the carrier frequency optimization.

where ω_c is the carrier frequency; $x_{HP}(t)$ and $x_{PV}(t)$ are the high-passed audio input and harmonics generated by the PV, respectively; $\hat{x(t)}_{HP}$ and $\hat{x_{PV}}(t)$ are their corresponding Hilbert transforms; "USB" and "LSB" are short for the upper and lower sidebands, respectively.

Generally, the frequency response of the ultrasonic emitter is analogous to a triangle [15]. Conventionally, the PAL sets the carrier frequency at the vertex of the triangle to maximize the out power of the carrier wave, *i.e.* $\omega_c = \omega_v$, where ω_c and ω_v denote the carrier frequency and the resonance frequency of the ultrasonic emitter, respectively.

A carrier frequency optimization is proposed to adjust the carrier frequency. As shown in Figure 2, when the carrier frequency is higher than the resonance frequency, the lower sideband can be emitted at higher efficiency. By doing so, the harmonics requested by the VB is preferably enhanced. However, the high-frequency components are attenuated due to the change of the carrier frequency. This will probably result in a decreased volume of the PAL. The optimal carrier frequency should be determined to balance the reproduction of both the harmonics for the VB and the high-passed audio input.

Therefore, the in-band peak flatness is proposed as a measure. The spectrum of the transmitted lower sideband $T_{LSB}(\omega|\omega_c)$ is written as

$$T_{LSB}(\omega|\omega_c) = H(\omega) \cdot Y_{LSB}(\omega|\omega_c), \tag{8}$$

where $H(\omega)$ and $Y_{LSB}(\omega|\omega_c)$ are the frequency response of the ultrasonic emitter and the spectrum of the lower sideband that carries the harmonics for the VB. Therefore, the in-band flatness F is calculated by

$$F(\omega_c) = -20\log_{10}\{\int_{\omega_c-B}^{\omega_c}[M - T_{LSB}(\omega)]d\omega\}, \tag{9}$$

where B is the bandwidth of the harmonics; and $M = max\{T_{LSB}(\omega)\}$ is the peak of the spectrum that has the greatest impact on listening experience.

The sound pressure level (SPL) of the PAL with respect to the carrier frequency is calculated by

$$SPL(\omega_c) = \frac{1}{N}\sum_{k=0}^{N}\{20\log_{10}(\frac{\sqrt{\frac{\sum_{i=1}^{M} t_k^2(i|\omega_c)}{M}}}{p_0})\} \tag{10}$$

where M is the length of the frame, typically given by 20 ms; N is the number of frames; $t_k(i|\omega_c)$ is the transmitted signal obtained by the inverse Fourier transform of $T(\omega|\omega_c)$ in the k-th frame; and $p_0 = 2 \times 10^{-5}$ Pa is the reference sound pressure. Moreover, $T(\omega|\omega_c)$ is the spectrum of the whole transmitted signal, which is written as

$$T(\omega|\omega_c) = H(\omega) \cdot [Y_{LSB}(\omega|\omega_c) + Y_{USB}(\omega|\omega_c)], \tag{11}$$

Table 1: List of music clips.

Title	Artist	Length	Genre
Ferry	Tsai Chin	8sec	Folk
Everybody	Backstreet Boys	10sec	Pop
The phantom of the opera	Andrew Webber	8sec	Opera

Table 2: List of processing methods.

Index	Method
1	High-pass filter (>400 Hz)
2	SSB modulation, without VB preprocessing
3	SSB modulation, with VB preprocessing
4	Proposed VSB modulation with VB preprocessing, without the carrier frequency optimization
5	Proposed VSB modulation with VB preprocessing, with the carrier frequency optimization
6	Unprocessed

where $Y_{USB}(\omega|\omega_c)$ is the spectrum of the upper sideband that embeds the high-passed audio input.

The optimal carrier frequency ω_o is determined by maximizing the in-band flatness with a SPL constraint, *i.e.*

$$\begin{aligned} \max_{\omega_c} \quad & F(\omega_c) \\ s.t. \quad & SPL(\omega_v) - SPL(\omega_c) \leq 10dB, \\ & \omega_c \geq \omega_v. \end{aligned} \tag{12}$$

Here, the SPL of the PAL is tolerated to be decreased by 10 dB.

3. SUBJECTIVE TEST AND RESULTS

A subjective test is conducted to evaluate the perception of the VB from the PAL. Three music clips with durations of 8-10 seconds are used in the subjective test. They are listed in Table 1. Six processing methods listed in Table 2 are compared by the multi-stimulus test with hidden reference and anchor (MUSARA) methodology. The unprocessed and high-passed audio input are treated as the hidden reference and anchor, respectively. The conventional SSB modulation is carried out with and without the PV-based VB preprocessing for comparison. The proposed VSB modulation is implemented with and without the carrier frequency optimization for comparison as well.

There are 20 participants aging from 15 to 25 years old, who are required to grade their perceived bass intensity and sound quality from 0 (poor) to 100 (excellent). The bass intensity refers to the perceived quality of the low-frequency sound, and the audio quality also takes noise and distortion into consideration. The participants have been briefed that the reference has standard scores of 50 for its bass intensity and 100 for its sound quality. The subjective test results are presented in Figure 3.

Based on the subjective test results, the sound quality of the PAL is basically unsatisfactory due to its severe harmonic distortion. The high-passed audio input serves the purpose of an anchor in the bass intensity but no in the sound quality. The results on the SSB modulation demonstrate that the perceived bass quality is greatly improved by the PV-based VB preprocessing. The effectiveness of the

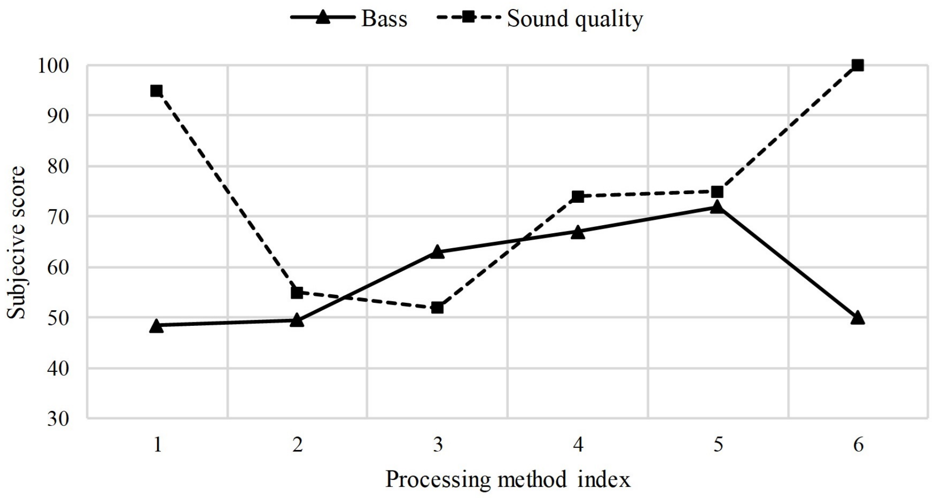

Figure 3: Subjective test results.

PV-based VB preprocessing is thus validated. The proposed VSB modulation outperforms the SSB modulation in both the bass quality and the sound quality. With the carrier frequency optimization, the bass quality of the VSB modulation can be further improved without degrading the sound quality.

4. CONCLUSIONS

In this paper, the VSB modulation is proposed to carry the high-passed audio input in the upper sideband and the harmonics generated for the VB in the lower sideband. It is noted that the lower sideband carrying the harmonics occupies a limited bandwidth. Therefore, the carrier frequency optimization method is proposed. The carrier frequency is no longer set at the resonance frequency of the ultrasonic emitter. Alternatively, it is offset to be higher than the resonance frequency in order for the harmonics in the lower sideband to be transmitted at higher power. The trade-off of changing the carrier frequency is the reduced volume of the PAL. Therefore, an objective measure, namely the in-band flatness, is introduced. It is maximized with the constraint of the SPL. The subjective test results prove that the PV-based VB preprocessing can effectively improve the perceived bass quality of the PAL and the proposed VSB modulation with the optimal carrier frequency results in the best overall listening experience.

5. ACKNOWLEDGEMENTS

This manuscript is prepared based on the research work supported by the National Natural Science Foundation of China and the Civil Aviation Administration of China (Joint Grant No. U1933127).

REFERENCES

[1] 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," *Journal of the Acoustical Society of America*, vol. 73, no. 5, pp. 1532-1536, 1983.

[2] P. J. Westervelt, "Parametric acoustic array," *Journal of the Acoustical Society of America*, vol. 35, no. 4, pp. 535-537, 1963.

[3] M. B. Bennett and D. T. Blackstock, "Parametric array in air," *Journal of the Acoustical Society of America*, vol. 57, no. 3, pp. 562-568, 1975.

[4] N. Tanaka and M. Tanaka, “Active noise control using a steerable parametric array loudspeaker,” *Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3526-3537, 2010.

[5] C. Shi, Y. Kajikawa, and W. S. Gan, “An overview of directivity control methods of the parametric array loudspeaker,” *APSIPA Transactions on Signal and Information Processing*, vol. 3, no. e20, pp. 1-12, 2014.

[6] C. Shi and Y. Kajikawa, “Volterra model of the parametric array loudspeaker operating at ultrasonic frequencies,” *Journal of the Acoustical Society of America*, vol. 140, no. 5, pp. 3643-3650, 2016.

[7] J. J. Croft and J. O. Norris, “Theory, history, and the advancement of parametric loudspeakers: A technology review,” American Technology Corporation, White paper, Part # 98-10006-1100 Rev.E, 2001.

[8] C. Shi, H. Mu and W. S. Gan, “A psychoacoustical preprocessing technique for virtual bass enhancement of the parametric loudspeaker,” *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, 2013.

[9] F. A. Karnapi, W. S. Gan and M. H. Er, “Method to enhance low frequency perception from a parametric array loudspeaker,” *Proceedings of 112th Audio Engineering Society Convention*, Munich, Germany, 2002.

[10] J. C. R. Licklider, “A duplex theory of pitch perception,” *Journal of the Acoustical Society of America*, vol. 23, no. 1, pp. 147-147, 1951.

[11] E. Larsen and R. M. Aarts, “Reproducing low-pitched signals through small loudspeakers,” *Journal of the Audio Engineering Society*, vol. 50, no. 3, pp. 147-164, 2002.

[12] N. Oo, W. S. Gan and M. O. J. Hawksford, “Perceptually-motivated objective grading of nonlinear processing in virtual-bass systems,” *Journal of the Audio Engineering Society*, vol. 59, no. 11, pp. 804-824, 2011.

[13] H. Mu, “Perceptual quality improvement and assessment for virtual bass system,” Ph. D. Thesis, Nanyang Technological University, 2015.

[14] C. Shi and Y. Kajikawa, “Automatic gain control for parametric array loudspeakers,” *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, Shanghai, China, 2016.

[15] C. Shi and Y. Kajikawa, “Effect of the ultrasonic emitter on the distortion performance of the parametric array loudspeaker," *Applied Acoustics*, vol. 112, pp. 108-115, 2016.