MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Virtual Bass Enhancement Based on Harmonics Control Using Missing Fundamental in Parametric Array Loudspeaker

GENG, Yuting[1]
Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga, Japan

NAKAYAMA, Masato[2]
Osaka Sangyo University / Ritsumeikan University
3-1-1 Nakagaito, Daito, Osaka, Japan / 1-1-1 Nojihigashi, Kusatsu, Shiga, Japan

NISHIURA, Takanobu[3]
Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga, Japan

ABSTRACT

Parametric array loudspeaker (PAL) can achieve a sharp directivity by using an ultrasonic wave. However, PAL has difficulty emphasizing bass sounds compared with conventional dynamic loudspeakers because it utilizes demodulation based on nonlinear interaction in the air. Also, the sound with PAL is degraded by harmonic distortion with intermodulation. Here, we focus on the missing fundamental, a psychoacoustical phenomenon indicating that human beings perceive a pitch of fundamental by harmonics. Based on this phenomenon, virtual bass enhancement has been proposed for small dynamic loudspeakers and applied on PAL. However, those harmonic generators are not suitable enough for PAL because PAL generates much more harmonic distortion than dynamic speakers. Therefore, in this paper, we propose virtual bass enhancement on basis of harmonics control adapted to harmonic distortion of PAL. The proposed method uses missing fundamental and employs a pre-processing to reproduce the virtual bass sound by stressing harmonics of acoustic signals. It re-allocates energy to higher harmonics to amplify the demodulated sound with high-shelf filter. In addition, the proposed method controls the harmonics to cause the missing fundamental and reduce the distortion of the demodulated sound using peak filters. Finally, we confirmed the effectiveness of the proposed method through evaluation experiments.

Keywords: Harmonics control, Virtual bass enhancement, Parametric array loudspeaker
I-INCE Classification of Subject Number: 00

[1] gr0369sv@ed.ritsumei.ac.jp
[2] nakayama@ise.osaka-sandai.ac.jp
[3] nishiura@is.ritsumei.ac.jp

1. INTRODUCTION

Audible sounds such as music and voice are usually reproduced by conventional dynamic loudspeakers. However, widely carried sounds become noise for non-listeners. In recent years, parametric array loudspeaker (PAL)[1,2] has received attention for its sharp directivity. PAL emits an intense amplitude modulated (AM) wave synthesized by modulating the ultrasonic wave with audible sound. The emitted intense AM wave is demodulated from ultrasonic wave to audible sound by nonlinear interaction in the air, and the demodulated audible wave achieves sharp directivity. Therefore, PAL can reproduce audible sound in a particular area[3], and is promising to be used in museums, stations and so on. However, there are two major limitations of the PAL which lead to a lower reproduction performance than conventional dynamic loudspeakers[2]. First, due to the physical principle of demodulation, it is difficult for PAL to reproduce bass sounds. Second, the sound quality is degraded by harmonic distortion and intermodulation. Since the poor bass reproduction of PAL is physically unsolvable[1], in order to reproduce the low-frequency sounds with a satisfying sound pressure, the total output energy should be increased significantly.

As an alternative solution, a psychoacoustic approach is carried out. Based on missing fundamental[4], the virtual bass enhancement generating harmonic series to let listeners perceive bass sounds is carried out for small-scale loudspeakers[5], and this study is also applied to the PAL[6]. Since PAL generates much more harmonic distortion over dynamic loudspeakers, the nonlinearity of PAL is considered. The conventional method adapted the harmonic generator for small-scale loudspeakers to the PAL and designed an equivalent harmonic generator to obtain the same harmonic components. The conventional method resulted in an improvement of bass sounds with a cost of severe sound quality losses. However, the conventional method is less effective for harmonic sounds (sounds with harmonic structure, e.g. voice, string instruments and so on) than inharmonic sounds (sounds without harmonic structure, e.g. drums and so on) and resulted in a smaller sound pressure improvement and bigger loss of sound quality.

Therefore, we focus on the difficulty of PAL in low-frequency reproduction and the harmonic distortion generated in demodulation. We suppress low-frequency components for an increasement of energy conversion efficiency. Missing fundamental is utilized for fundamental perception with harmonic series, so the energy of low-frequency components can be allocated to other components for an increasement of energy conversion efficiency. Double sideband (DSB) modulation can achieve a higher sound pressure over other modulation method but generates more harmonic distortions. We make use of these harmonic distortions, which are regarded as useless by-products in past studies. First, frequencies of harmonic distortions are multiples of fundamental, so they contribute to create missing fundamental effect. Moreover, harmonic distortions generated by low-frequency components have a same frequency as high-frequency components in audible sound, so we can utilize harmonic distortions to enhance high-frequency components.

In this paper, we propose a virtual bass enhancement based on harmonics control for harmonic sounds with low fundamental frequency. We utilize high shelving filter (HSF) to suppress the low-frequency components, allocate energy to high-frequency components, and improve the sound pressure of demodulated sound. Missing fundamental is created in a higher sound pressure as a compensation of sound quality to suppression of low-frequency components. Then, we utilize a sequence of peaking filters (PKF) to control the proportion of harmonic components under a designed strategy using the inverse frequency response of PAL. Due to the harmonics control, the loss of sound quality can be reduced, and an improvement of sound quality is possible to achieve.

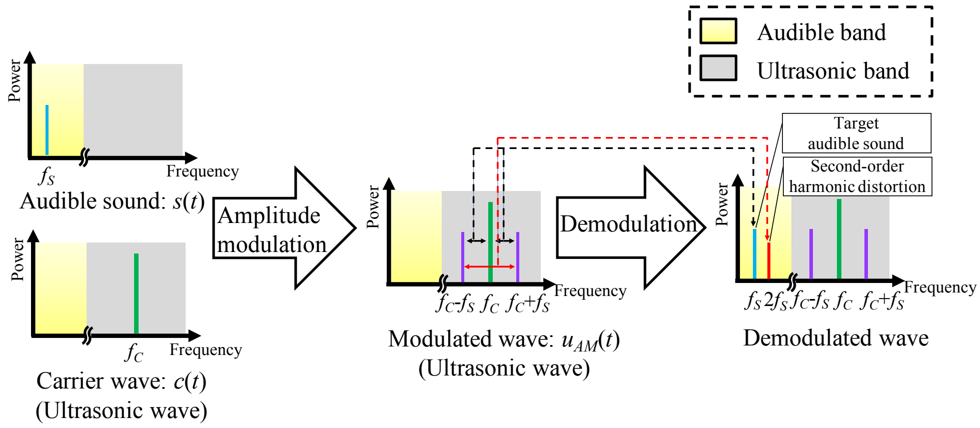


Figure 1 Overview of modulation and demodulation of parametric array loudspeaker

2. PARAMETRIC ARRAY LOUDSPEAKER

2.1 Principle of Parametric Array Loudspeaker

PAL can realize a sharper directivity by utilizing a parametric array effect[1] in the air. Figure 1 shows the overview of modulation and demodulation of PAL. PAL emits an intense amplitude modulated (AM) wave which is synthesized by modulating an ultrasonic carrier wave with audible sound. The carrier wave $c(t)$ and audible sound $s(t)$ can be indicated as follows:

$$c(t) = A_C \cos(2\pi f_C t), \quad (1)$$

$$s(t) = A_S \cos(2\pi f_S t), \quad (2)$$

where t denotes time index, A_C and A_S denote the amplitudes of carrier wave and audible sound, f_C and f_S denote the frequencies of carrier wave and audible sound respectively. The emitted AM wave $u_{AM}(t)$ can be indicated as follows:

$$u_{AM}(t) = \{1 + ms(t)\}c(t), \quad (3)$$

where $m = A_S/A_C$ $(0 < m \leq 1)$ denotes the modulation factor. Equation (3) can be transformed as follows with Eqs. (1) and (2):

$$u_{AM}(t) = A_C \cos(2\pi f_C t) + \frac{1}{2}{A_S}^2\{\cos(2\pi(f_C + f_S)t) + \cos(2\pi(f_C - f_S)t)\}. \quad (4)$$

From Eq. (4), the AM wave consists of carrier (frequency: f_C), upper sideband (frequency: $f_C + f_S$), and lower sideband (frequency: $f_C - f_S$). The emitted AM wave is self-demodulated into audible sound by nonlinear interaction in the air. The demodulated audible sound is the difference in frequency domain between carrier wave and upper sideband, or between carrier wave and lower sideband. Since the target sound is a result of interaction between two ultrasonic waves, it derives the sharp directivity of ultrasonic waves. Thus, PAL can carry audible sound to particular area.

2.2 Problems of Parametric Array Loudspeaker

There are two major problems with the PAL. One is loss of sound quality due to harmonic distortion, and the other one is low sound pressure in low-frequency band. First, PAL suffers from harmonic distortions much more than conventional dynamic loudspeakers since it utilizes the nonlinear interaction in the air. As shown in Fig. 1, in case of utilizing the DSB modulation, the original audible wave is demodulated as the difference between carrier wave and each sideband. However, the second harmonic distortion is demodulated between two sidebands. Intermodulation will also occur much more if the original audible sound has several frequency components. The harmonic distortions and intermodulation cause the sound quality degradation of PAL. In addition,

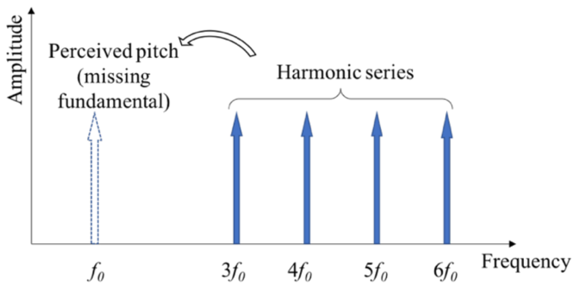


Figure 2 Missing fundamental effect

PAL has theoretical difficulties in reproducing bass sounds[1]. The nonlinearity of the PAL can be described with the sound pressure of demodulated sound as follows:

$$p \propto \frac{p_0}{d}\frac{\partial^2}{\partial t'^2}|E(t')|, \tag{5}$$

where p denotes the sound pressure of demodulated wave at observation point on the direction of propagation, p_0 denotes the source sound pressure, d denotes the distance between PAL and observation point, $t' = t - d/c_0$ is the retarded time with speed of sound c_0, $E(t')$ is the envelope of AM wave. In the case when $s(t)$ is a sinusoidal wave as indicated in Eq. (2), Eq. (5) can be altered as

$$p \propto \frac{p_0}{d} \cdot 4\pi^2 f_S^2 m A_S \cos(2\pi f_S t'). \tag{6}$$

Equation (6) suggests that the sound pressure of demodulated wave is in direct proportion to the square of the frequency of audible sound. Owing to this relationship, bass sounds are reproduced by PAL with an extremely low sound pressure.

3. MISSING FUNDAMENTAL

Voices and sounds of instrument consist of fundamental (frequency: f_0) and harmonic components (frequency: $2f_0, 3f_0, 4f_0, \dots, Nf_0$). This structure in frequency is called harmonic structure[7]. When listening to these harmonic sounds, human beings perceive a frequency of fundamental. Pitch is defined as a subjective psychophysical quantity which describes the frequency perception by human auditory system. As shown in Fig. 2, even if the fundamental does not physically exist, pitch of the "missing" fundamental can still be perceived. This phenomenon is called "missing fundamental[4]". The existence of missing fundamental implies that it is possible to create a low-frequency perception with high-frequency components. And this possibility can be realized by generating harmonic series of target sound.

4. CONVENTIONAL VIRTUAL BASS ENHANCEMENT BASED ON HARMONIC GENERATION FOR PARAMETRIC ARRAY LOUDSPEAKER

Missing fundamental opens up a psychoacoustic approach for bass enhancement. Nonlinear device or nonlinear function are used to generate harmonic series for bass sounds, so that loudspeakers with poor bass quality can reproduce "virtual" bass sounds. And this virtual bass enhancement is applied on PAL[6]. Figure 3 shows the block diagram of conventional method. The audio input $s(t)$ is split into lower-frequency sound $s_L(t)$ and higher-frequency sound $s_H(t)$. $s_L(t)$ and $s_H(t)$ are indicated as follows:

$$s_L(t) = s(t) * h_{LPF}(t), \tag{7}$$

$$s_H(t) = s(t) * h_{HPF}(t), \tag{8}$$

where $h_{LPF}(t)$ and $h_{HPF}(t)$ denote the low-pass filter and high-pass filter. The cut-off frequency of each filter is 500 Hz. The lower-frequency sound $s_L(t)$ is processed by a harmonic generator to create missing fundamental by harmonic series of low-frequency

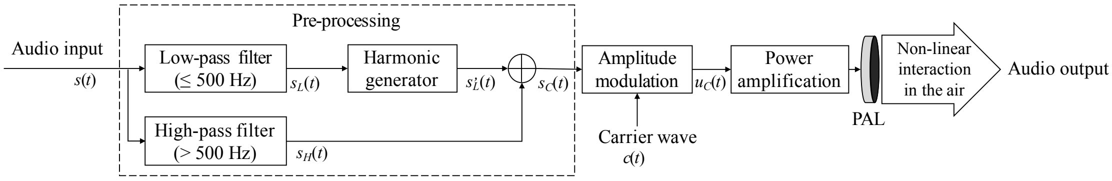


Figure 3 Block diagram of the conventional method

components. The nonlinear function $F_{ATSR}(x)$ is used in harmonic generator and it is indicated as follows:

$$\begin{aligned} F_{ATSR}(x) = &-0.968x^6 + 0.667x^5 + 0.783x^4 \\ &-0.184x^3 - 0.086x^2 + 0.383x + 0.361, \end{aligned} \tag{9}$$

where x denotes the normalized value of signal ($-1 \leq x \leq 1$). This function is an altered form of arc-tangent square root (ATSR) function[8] which is utilized in virtual bass enhancement. Since more harmonic distortion occurs during the reproduction of PAL than that of conventional dynamic loudspeakers, in order to generate same harmonic series for PAL as what generated by ATSR for conventional dynamic loudspeakers, the nonlinearity of PAL[1] is considered. As a result, $F_{ATSR}(x)$ is calculated as shown in Eq. (9). The processed lower-frequency sound $s_L'(t)$ is then added to high-frequency sound. The output of pre-processing $s_C(t)$ and modulated wave $u_C(t)$ are indicated as:

$$s_C(t) = s_L'(t) + s_H(t), \tag{10}$$

$$u_C(t) = \{1 + m s_C(t)\} c(t). \tag{11}$$

After emitting the modulated wave in the air, missing fundamental is created in demodulated sound, and listener can perceive bass sounds with higher sound pressure.

However, the effectiveness of conventional method is limited when processing harmonic sounds: the conventional method resulted in an insufficient sound pressure and more loss of sound quality. First, because low-frequency components are kept after generating harmonic series for them, the energy conversion efficiency is still low, and the sound pressure level is not satisfying. Meanwhile, adding low-frequency sounds on high-frequency sounds leads to clipping distortion and causes the loss of sound quality. Moreover, not only fundamental but also some low-frequency harmonic components might be processed. Harmonic series generated for these components resulted in more harmonic distortions. Finally, in the conventional method, intermodulation between harmonic components is not considered in the nonlinear model of PAL, as a result, harmonic distortions occur over expectation in reproduction.

5. PROPOSED VIRTUAL BASS ENHANCEMENT BASED ON HARMONIC CONTROL FOR PARAMETRIC ARRAY LOUDSPEAKER

5.1 Overview of Proposed method

To solve the problems of conventional method when processing harmonic sounds, we focus on the harmonic structure. It makes no sense to generate harmonic series for harmonic sounds, because the harmonic series generated for low- frequency components have same frequency as some existing components. Hence, we control the proportion of harmonic components to realize an improvement of sound pressure and sound quality. The low-frequency components are difficult for PAL to reproduce so they are suppressed to allocate energy to components which are easy for PAL to reproduce. Owing to this suppression, the energy conversion efficiency increases, and the total sound pressure level is improved. Moreover, a pitch of fundamental can be still perceived with the occurrence of missing fundamental though the fundamental is suppressed. On the other hand, Harmonic distortions are regarded as useless by-products and are targets to be

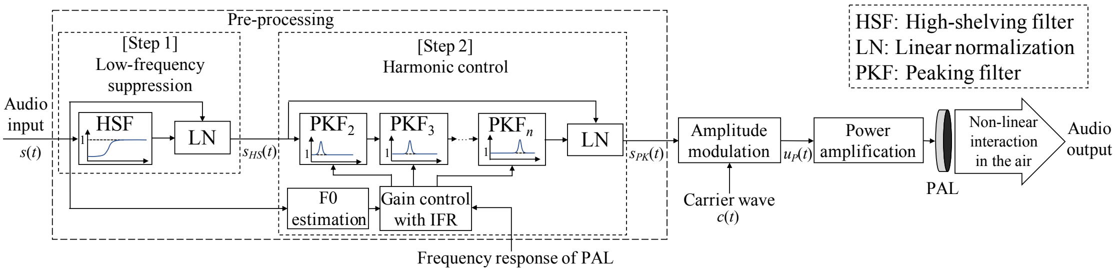


Figure 4 Block diagram of the proposed method

eliminated in past studies, however, we try to make use of these harmonic distortions. By considering the occurrence of harmonic distortions and controlling the proportion of harmonic components in advance, it is possible to improve the sound quality.

In this paper, we propose a virtual bass enhancement based on harmonics control for PAL. Figure 4 shows the block diagram of the proposed method. The proposed method consists of two steps: Step 1 is a low-frequency suppression to improve the sound pressure, Step 2 is a harmonics control to improve the sound quality. In Step 1 of the proposed method, HSF is used to suppress the fundamental and low-frequency harmonic components. After the suppression, a linear normalization (LN) is operated to keep the amplitude of signal same as that before processed. With this LN, more energy is allocated to high-frequency components so that the energy conversion efficiency is improved, and the sound pressure of demodulated sound can be improved. The input of Step 1 is the audio input $s(t)$, and the processed signal is denoted by $s_{HS}(t)$. In Step 2 of the proposed method, a sequence of PKFs is used to control the proportion of harmonic components. To determine the parameters of PKFs, F0 estimation[9] is added to estimate the fundamental frequency. Gain control strategy using inverse frequency response (IFR) of PAL is also designed to determine the gain of each PKF in order to improve the sound quality of demodulated sound. After the process by PKFs, a LN is also operated to keep the amplitude. Signal processed by Step 1 $s_{HS}(t)$ is input to Step 2, and the output of Step 2 is denoted by $s_{PK}(t)$. After processed by Steps 1 and 2 in proposed method, the audible signal $s_{PK}(t)$ is modulated into ultrasonic wave $u_P(t)$ and propagated through PAL.

5.2 [Step 1] Low-frequency Suppression Using High Shelving Filter

Figure 5 shows the process of Step 1 in the proposed method. For a higher energy conversion efficiency, components lower than cutoff frequency f_{HS} is suppressed to allocate more energy to higher components. With amplification of higher components, the sound pressure of demodulated sounds increases, and more harmonic distortions are generated during demodulation. A higher sound pressure is achieved, and the pitch is maintained. HSF can suppress low-frequency components and enhance high-frequency components. The transfer function of a Biquad HSF[10] $H_{HSF}(z)$ can be indicated as:

$$H_{HSF}(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{a_0 + a_1 z^{-1} + a_2 z^{-2}}. \quad (12)$$

And the coefficients are indicated as follows:

$$\begin{cases} b_0 = \quad A\{A + 1 + (A-1)\cos(\omega_{HS}) + 2\sqrt{A}\alpha\} \\ b_1 = -2A\{A - 1 + (A-1)\cos(\omega_{HS})\} \\ b_2 = \quad A\{A + 1 + (A-1)\cos(\omega_{HS}) - 2\sqrt{A}\alpha\} \end{cases},$$
$$\begin{cases} a_0 = \quad A + 1 - (A-1)\cos(\omega_{HS}) + 2\sqrt{A}\alpha \\ a_1 = 2\{A - 1 + (A-1)\cos(\omega_{HS})\} \\ a_2 = \quad A + 1 - (A-1)\cos(\omega_{HS}) - 2\sqrt{A}\alpha \end{cases}, \quad (13)$$

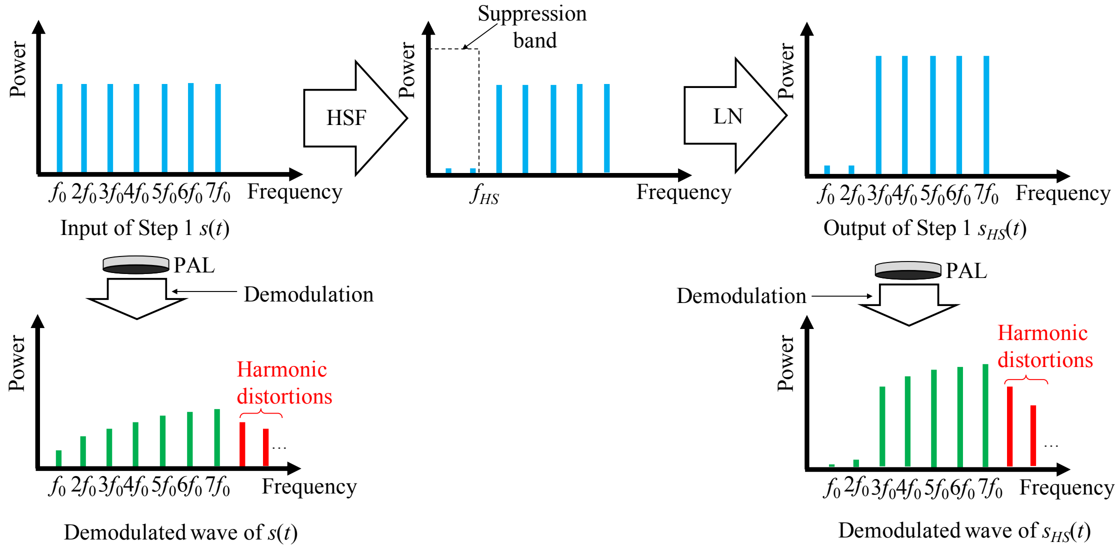


Figure 5 Process of Step 1 in the proposed method

where $\omega_{HS} = 2\pi f_{HS}/F_S$ denotes the cutoff frequency f_{HS} in radian, F_S [Hz] denotes the sampling frequency of original signal. Intermediate variables A and α are indicated as:

$$A = 10^{G_{HS}/40}, \tag{14}$$

$$\alpha = \frac{1}{2}\sin(\omega_{HS}) \cdot \sqrt{(A + A^{-1})(S^{-1} - 1) + 2}, \tag{15}$$

where G_{HS} [dB] denotes the gain of HSF, S [dB/oct] denotes the slope parameter.

In Step 1, after low-frequency suppression on original signal $s(t)$, LN is carried out, so that the output $s_{HS}(t)$ has the same amplitude as $s(t)$. By returning amplitude to former value, unsuppressed components are enhanced, and the allocation of energy is realized. The output of Step 1 $s_{HS}(t)$ can be indicated as follows:

$$s_{HS}(t) = \mathbb{N}_{s(t)}[s(t) * h_{HSF}(t)], \tag{16}$$

where $h_{HSF}(t)$ denotes the transfer function of HSF $H_{HSF}(z)$ in time domain, $*$ denotes the convolution operator, $\mathbb{N}$ denotes the LN process. In Eq. (16), signal process by HSF is normalized to have the same maximum amplitude as $s(t)$. The LN process which normalize signal $b(t)$ with signal $a(t)$ can be indicated as follows:

$$\mathbb{N}_{a(t)}[b(t)] = b(t) \cdot \frac{\max_t[|a(t)|]}{\max_t[|b(t)|]}, \tag{17}$$

where max denotes the function to derive the maximum, $\max_t[|a(t)|]$ denotes the maximum of $|a(t)|$ in time domain. Owing to the maintenance of amplitude, energy consumption of reproduction varies little, and clipping distortions can be avoided.

5.3 [Step 2] Harmonics Control Using Peaking Filter

Figure 6 shows the process of Step 2 in the proposed method. In Step 2, we control the proportion of harmonic components to compensate the degraded sound quality and aim at an improvement on sound quality. Here, we notice that harmonic distortion generated by lower harmonic components has a same frequency as some higher harmonic components of the original sound. Therefore, we amplify some lower components and reduce the energy of higher components in original sound. Higher components are enhanced by harmonic distortions generated from lower ones in demodulation. And harmonic distortions generated from these components can be limited. The sound quality is improved as a result, and energy conversion efficiency can be furtherly improved to achieve a higher sound pressure. PKF can amplify components around certain frequency. We utilize a sequence of PKFs to control each harmonic component. F0 estimation[9]

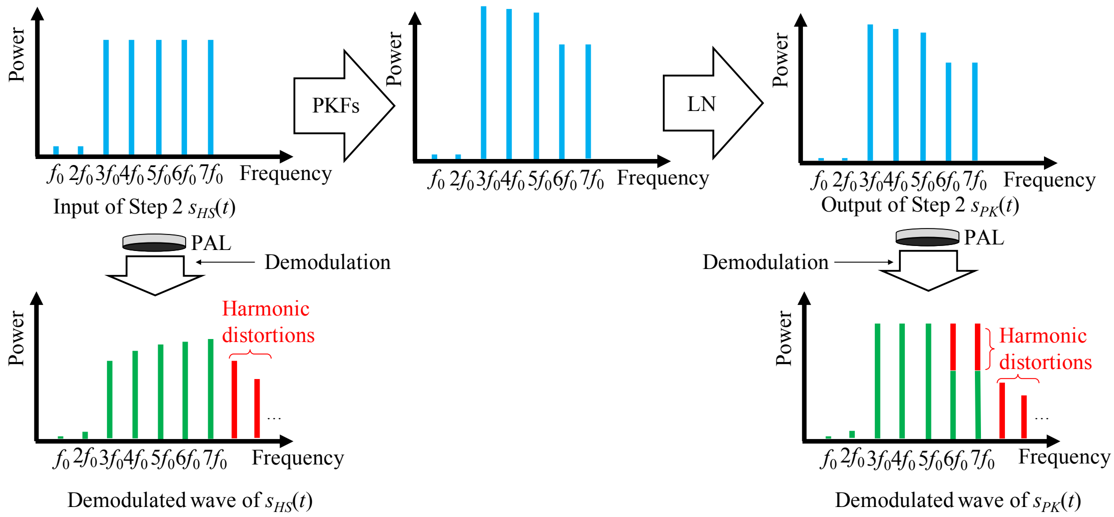


Figure 6 Process of Step 2 in the proposed method

is needed before designing PKFs to calculate the frequency of each harmonic component. The transfer function of Biquad PKF[10] controlling the n-th harmonic component $H_{PKF_n}(z)$ can be indicated as:

$$H_{PKF_n}(z) = \frac{d_{0,n} + d_{1,n}z^{-1} + d_{2,n}z^{-2}}{c_{0,n} + c_{1,n}z^{-1} + c_{2,n}z^{-2}}. \tag{18}$$

And the coefficients are indicated as follows:

$$\begin{cases} d_{0,n} = 1 + \beta_n B_n \\ d_{1,n} = -2\cos(\omega_{PK_n}) \\ d_{2,n} = 1 - \beta_n B_n \end{cases}, \begin{cases} c_{0,n} = 1 + \beta_n B_n^{-1} \\ c_{1,n} = -2\cos(\omega_{PK_n}), \\ c_{2,n} = 1 - \beta_n B_n^{-1} \end{cases} \tag{19}$$

where $\omega_{PK_n} = 2\pi n f_0/F_S$ denotes the center frequency of PKF controlling the n-th harmonic component in radian, f_0 denotes the fundamental frequency of original signal which is measured by F0 estimation. Intermediate variables β_n and B_n are indicated as:

$$\beta_n = \sin(\omega_{PK_n}) \cdot \sinh\left(\frac{\log_e(2b\omega_{PK_n})}{2\sin(\omega_{PK_n})}\right), \tag{20}$$

$$B_n = 10^{G_{PK_n}/40}, \tag{21}$$

where b [oct] denotes the width of peak, G_{PK_n} [dB] denotes the gain of PKF$_n$ determined by the gain control with IFR. The gain control strategy will be discussed in Sec. 5.4.

A LN similar as that indicated in Eq. (17) is also carried out after harmonics control. The output of Step 2 $s_{PK}(t)$ can be indicated as follows:

$$s_{PK}(t) = \mathbb{N}_{s_{PK}(t)}[s_{PK}(t) * h_{PKF}(t)], \tag{22}$$

$$h_{PKF}(t) = h_{PKF_2}(t) * \cdots * h_{PKF_n}(t) * \cdots * h_{PKF_N}(t), \tag{23}$$

where $h_{PKF_n}(t)$ denotes the transfer function of PKF$_n$ $H_{PKF_n}(z)$ in time domain. Pre-processed sound $s_{PK}(t)$ is then used in AM and the emitted AM wave is indicated as:

$$u_P(t) = \{1 + ms_{PK}(t)\}c(t). \tag{24}$$

The modulated wave is then emitted by PAL to reproduce audible sound.

5.4 Gain Control of Peaking Filter with Inverse Frequency Response

In Step 2, it is ideal to set the gain of PKFs under the variation of harmonic structure. However, the demodulation is complex due to the intermodulation and harmonic distortions, which makes it difficult to design the best gain control strategy. As a compromise, we propose a gain control strategy using the IFR of PAL. The degradation of sound quality caused by PAL's non-flat frequency response (FR) can be compensated.

The calculation of IFR can be described as:

$$I(f) = \frac{\max_f \left[|R(f)|\right]}{|R(f)|}, \tag{25}$$

where $I(f)$ denotes the IFR of PAL, $R(f)$ denotes the amplitude spectrum of PAL which represents the FR of PAL, f [Hz] denotes the frequency index ($0 < f < 20000$ [Hz]), $\max_f \left[|R(f)|\right]$ denotes the maximum of $|R(f)|$ in frequency domain. The maximum of FR is introduced to guarantee that the value of G_{PK_n} is bigger than 0. The FR of PAL should be measured in advance and under the same conditions as reproduction. Then the gain of PKFs in Step 2 can be set with the value of IFR at corresponding frequency. Under this gain control, G_{PK_n} in Sec. 5.3 can be indicated as follows:

$$G_{PK_n} = \begin{cases} 20\log_{10}\{I(nf_0)\}, & \text{if } n_L < n \le n_U \\ 0, & \text{otherwise} \end{cases}. \tag{26}$$

Harmonic components suppressed in Step 1 will not be amplified, so we set the lower limit $n_L = f_{HS}/f_0$. As we discussed in Sec. 5.3, harmonic components with high frequency will not be amplified in advance, so we set the upper limit $n_U = N/2$, where N denotes the total number of harmonic components existing in audio input.

6. EVALUATION EXPERIMENT FOR SOUND PRESSURE AND SOUND QUALITY

6.1 Experimental Conditions

We carried out evaluation experiments to confirm the effectiveness of the proposed method by sound pressure and sound quality. The demodulated sounds of PAL are recorded under conditions shown in Table 1 and with equipment shown in Table 2. The sound pressure is evaluated by sound pressure level (SPL), and is calculated as:

$$SPL_{obj} = SPL_{ref} + 10\log_{10}\frac{\sigma_{obj}^2}{\sigma_{ref}^2}, \tag{27}$$

where SPL_{obj} [dB] denotes the SPL of object sound in audible band (0~20000 Hz), SPL_{ref} [dB] denotes the SPL of reference sound with calibration of microphone, σ_{obj} and σ_{ref} denote the standard deviation calculated from time waveform of object sound and reference sound. A higher value of SPL suggests a better performance.

The sound quality is evaluated by spectral distortion (SD), and is calculated as:

$$SD = 10\log_{10}\sqrt{\frac{1}{F}\sum_{f=1}^{F}|S(f) - X(f)|^2}, \tag{28}$$

where f [Hz] denotes a frequency index, F [Hz] denotes the maximum of evaluation band and is set to 20000 Hz in this evaluation, $S(f)$ denotes a normalized power spectrum of original audible sound, $X(f)$ denotes a normalized power spectrum of demodulated sound. SD indicates the difference of power spectrum between object sound and reference sound, so a lower value suggests better sound quality.

In evaluation experiments, we used samples in RWC Music Database[12] as sound source, since sounds of instrument have an obvious harmonic structure and a steady frequency in a certain note. The fundamental frequency is known, so F0 estimation can be omitted in evaluation experiments to avoid the experimental results being influenced by the performance of F0 estimation. Both conventional and proposed methods affect more on sounds with lower fundamental than ones with higher fundamental. So, we chose a sound source of contrabass with a fundamental of 110 Hz which is extremely difficult for PAL to reproduce, and a sound source of guitar with a fundamental of 440 Hz which

Table 1 Experimental conditions

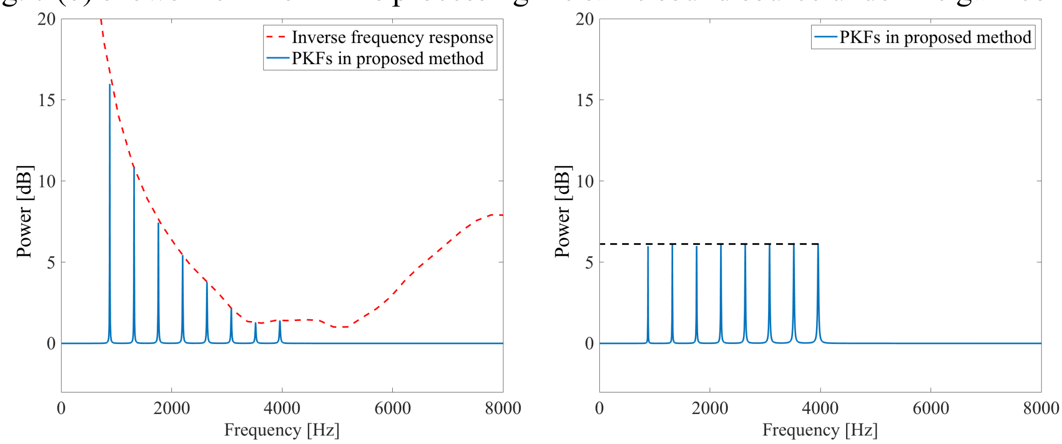
Ambient noise level	30.8 dB (A-weighted)
Reverberation time	$T_{[60]}$ =650 ms
Distance between PAL and microphone	2.0 m
Sampling frequency	192 kHz
Quantization bit rate	16 bits (reproduction) 32 bits (recording)
Sound source	RWC Music Database: Musical Instrument Sound[12] Sound of Contrabass ($f_0 = 110$ Hz) Sound of Guitar ($f_0 = 440$ Hz)
Modulation	DSB ($f_C = 40$ kHz and $m = 1.0$)
Gain of HSF	$G_{HS} = -30$ dB (Step 1 of proposed method)
Cut-off frequency of HSF	$f_{HS} = 500$ Hz (Step 1 of proposed method)

Table 2 Experimental equipment

Parametric array loudspeaker	MITSUBISHI, MSP-50E
Power amplifier	VICTOR, PS-A2002
A/D, D/A converter	RME, FIREFACE UFX
Microphone	SENNHEISER, MKH416

is relatively easy for PAL to reproduce.

As a comparison, we prepared five pre-processing methods for each sound source. First, non-processed sound is prepared as a reference to show the effect of each method. The second one is processed by conventional method[6].To confirm the effectiveness of Step 1 of proposed method, we prepared a sound processed only by HSF. Moreover, to confirm the effectiveness of gain control strategy in Step 2 of proposed method, we prepared a pair of sounds processed by Steps 1 and 2 but with different gain control strategies in Step 2: one is the gain control with IFR of PAL described in Sec. 4.3; the other one is gain control without IFR, which amplifies each component with even gain. We first carried out a preliminary experiment to measure the FR of PAL with time-stretched pulse (TSP) signal[11]. The conditions and equipment are same as that of evaluation experiments. The calculated IFR of PAL is shown in Fig. 7 (a) with the FR of PKFs processing sound source of guitar under the gain control proposed in Sec. 5.4. And Fig. 7 (b) shows the FR of PKFs processing the same sound source under the gain control



(a) Gain control with IFR *(b) Gain control without IFR*

Figure 7 The frequency response of PKFs with different gain control strategies in Step 2 of proposed method for processing the guitar sound sample

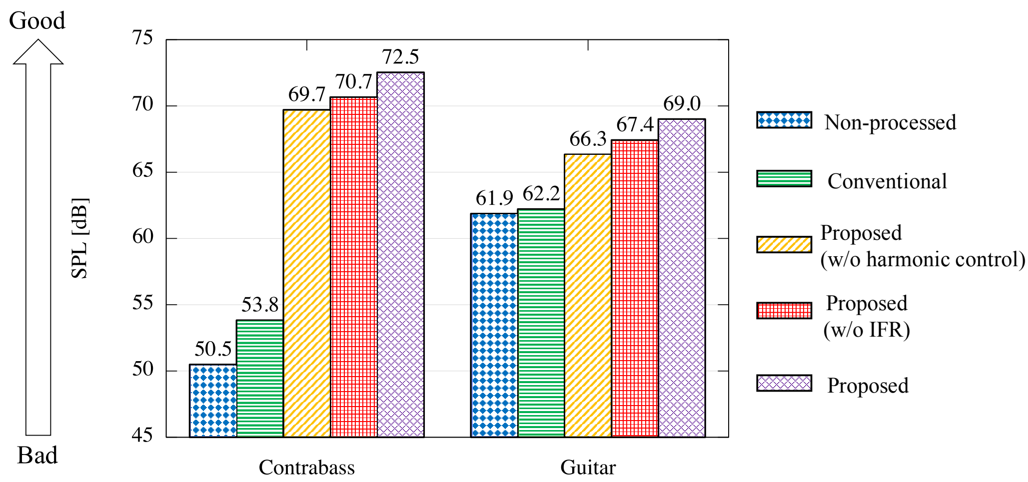


Figure 8 Experimental result on sound pressure

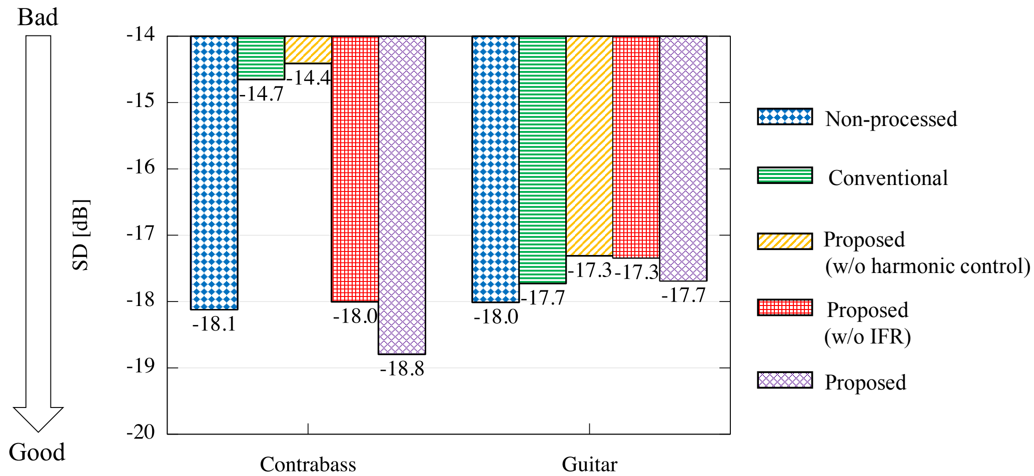


Figure 9 Experimental result on sound quality

without IFR. Each gain of PKF in Fig. 7 (b) is calculated by taking average of all gains in Fig. 7 (a) for each component. Each sound source is reproduced and recorded 32 times to calculate an average SPL and SD.

6.2 Experimental Results

The experimental results on sound pressure and sound quality are shown in Figs. 8 and 9 respectively. First, it is observed that conventional method achieves an improvement on sound pressure and causes losses of sound quality by comparing "Conventional" with "Non-processed". It is also observed that the HSF in Step 1 of proposed method degrades the sound quality but achieves a sufficient sound pressure improvement by comparing "Proposed (w/o harmonic control)" with "Conventional". This result proves the effectiveness of Step 1 in proposed method. Then, comparing "Proposed" with "Proposed (w/o harmonic control)", the result shows that sound quality is improved by Step 2 of proposed method and sound pressure is furtherly improved. Moreover, comparing "Proposed" with "Proposed (w/o IFR)", it is observed that our gain control strategy using IFR of PAL shows a better performance than that using even gain. These results prove the effectiveness of Step 2 in proposed method and imply that a better sound quality is promising with a well-optimized gain control strategy of harmonics control. In addition, by comparing "Proposed" with "Conventional", we confirm that proposed method has a better performance over conventional method, since it achieves a more sufficient sound pressure improvement and a less loss of sound quality. Comparing

"Proposed" with "Non-processed", it is observed that the sound quality of guitar is degraded but with a very little loss, and the sound quality of contrabass is improved to a better level by proposed method. This result implies that sound pressure and sound quality can be improved in the same time with an effective pre-processing method. As a conclusion of evaluation experiments, the proposed method achieves a sufficient improvement of sound pressure with a better sound quality.

7. CONCLUSION

In this paper, we proposed a new virtual bass enhancement based on harmonics control for PAL to reproduce low-frequency harmonic sound in a higher sound pressure with an equal or better sound quality. By evaluation experiment, we confirm that proposed method can improve the sound pressure and maintain the sound quality though some low-frequency components are suppressed. The experimental result also reveals possibility on a further improvement of sound quality. In future, it is hoped to derive the variation law of harmonic structure in demodulation of PAL and design a more effective harmonics control method for PAL.

8. ACKNOWLEDGEMENTS

This work was partly supported by Ritsumeikan R-GIRO, JST COI, JST SCORE and JSPS KAKENHI Grant Numbers JP18K19829, JP18K11365.

9. REFERENCES

1. P. J. Westervelt, "Parametric acoustic array", *J. Acoust. Soci. Am.*, 35(4), 535-537, (1963).
2. K. Aoki, T. Kamakura, Y. Kumamoto, "Parametric loudspeaker–characteristics of acoustic field and suitable modulation of carrier ultrasound", *Electronics and Communications in Japan*, 74(9), 76-82, (1991).
3. M. Yoneyama, J. Fujimoto, K. Kawamo, S. Sasabe, "The audio spotlight: An application of nonlinear interaction of sounds to a new type of loudspeaker design", *J. Acoust. Soci. Am.*, 73(5), 1532-1536, (1983).
4. J. C. R. Licklider, "A duplex theory of pitch perception," The Journal of the Acoustical Society of America, 23(1), 147-147, (1951).
5. E. Larsen, R. M. Aarts, "Reproducing low-pitched signals through small loudspeakers", *Journal of the Audio Engineering Society*, 50(3), 147–164, (2002).
6. C. Shi, H. Mu, W.S. Gan, "A psychoacoustical preprocessing technique for virtual bass enhancement of the parametric loudspeaker," *ICASSP2013,* 31-35, (2013).
7. J. Bharucha, C. L. Krumhansl, "The representation of harmonic structure in music: Hierarchies of stability as a function of context," *Cognition*, 13(1), 63-102 (1983).
8. N. Oo, W. S. Gan, W. S. Lim, "Generalized harmonic analysis of arc-tangent square root (ATSR) nonlinear device for virtual bass system," *ICASSP2010*, 301-304, (2010).
9. M. Goto, A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*, 43(4), 311-329, (2004).
10. U. Zölzer, "Digital Audio Signal Processing", New York: John Wiley and Sons, Inc., (1999).
11. G. B. Stan, J. J. Embrechts, D. Archambeau, "Comparison of different impulse response measurement techniques", *Journal of the Audio Engineering Society*, 50(4), 249-262, (2002).
12. M. Goto, "Development of the RWC music database." *Proceedings of the 18th International Congress on Acoustics*, 1, 553-556, (2004).