A hybrid bass enhancement system with improved transient and steady-state performance

Last name of 1st Author, first name of 1st Author, Last name of 2nd Author, First name of 2nd Author…
Affilication1: dept. name of organization, name of organization, City, Country
Affilication2: dept. name of organization, name of organization, City, Country
e-mail of 1st Author, e-mail of 2nd Author, …

Abstract—**Bass effect is an important criterion for audio system. However, the small loudspeakers in portable devices have poor low frequency responses. Conventional methods to enhance the bass effect using equalizers does not help significantly and may result in distortion or permanent damage to the loudspeakers. Recently, the virtual bass system (VBS) based on the psychoacoustic phenomenon called "missing fundamental" has been proposed, whereby human auditory system can perceive the fundamental frequency from its higher harmonics. Nonlinear devices (NLD) and phase vocoder (PV) are commonly used to generate harmonics in VBS. Yet, both approaches have their strength and weakness, so the hybrid VBS, combined the two approaches together, has been proposed to overcome the drawbacks. The conventional hybrid VBS needs a transient content detector to determine the weight of NLD and PV process, which has a poor transient and steady-state performance. In this paper, we proposed a new hybrid VBS using the harmonic and percussive sound separation (HPSS) algorithm to process the music signal through NLD and PV separately. To improve the bass effect as well as maintain the audio quality, we utilize the improved PV and subjective based harmonic magnitude control (HMC). Experiments show that the bass effect has been improved significantly with maintaining the good audio quality and the processing efficiency has increased too.**

Index Terms—virtual bass; bass enhancement; HPSS; audio quality; psychoacoustics

I. INTRODUCTION

As multimedia devices are getting smaller, thinner and lighter, loudspeaker units that are embedded in these devices must be reduced in size and thickness. However, the problem of the small-sized loudspeakers is that they cannot produced good bass (low frequency) effect due to their cost constraint and physical size limitation. But the bass effect is essential for listening feeling. The conventional method to dealing with this problem is to use the equalizers. The low frequency power is increased by shelve filters and other electronic means, which always lead to poor bass effect. Furthermore, direct bass boosting will result in nonnliner distortions, or even permanent damage to the loudspeaker.

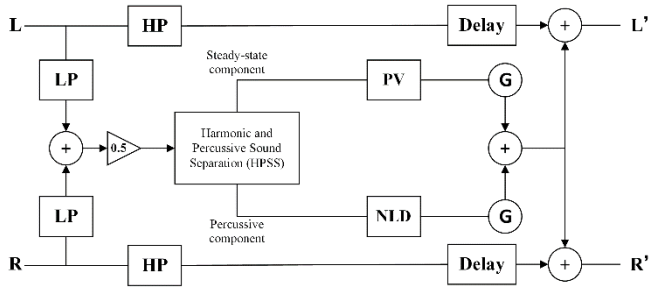
In previous study [1], a psychoacoustics phenomenon called "missing fundamental" was investigated and it can be used to generate virtual bass effect. The missing fundamental implies that the higher harmonics of the fundamental frequency can produce the presence of the fundamental frequency in the human auditory system. Almost all virtual bass algorithms utilize a nonlinear devices (NLD) or a phase vocoder (PV) to generate harmonics. NLD approach [2-5] are generally used in the time domain and various nonlinear devices come into use. MaxxBass [6] is the first virtual bass system (VBS) which has been on the market and it utilizes a multiplication circuit to generate harmonics. The NLD approach operates in the time domain and may introduces harmonic distortion according to the specific nonlinear devices chosen. This method produces good subjective listening result for percussive sound, such as drum beats. However, NLD introduces intermodulation distortion and does not allow for accurate control over the specific harmonic components, which will lead to an unnatural virtual bass effect for pitched signal components.

In the frequency domain, PV approach allows for accurate control over the individual harmonic components. Bai [8] proposed a virtual bass system based on PV and it uses pitch shifting to generate harmonic components. Compare with the NLD, the PV approach eliminates the problem of intermodulation by precisely controlling the individual harmonic components. But the PV approach needs a sufficient large analysis window in the time domain to achieve enough low-frequency resolution, which will lead to a smearing of percussive signal component. And it will introduce unnatural effects in musical signals, so PV is more applicable for pitched or stead-state signals than NLD.

To take the advantages of both NLD and PV, the hybrid virtual bass system [9] has been proposed recently. In Hill's method, NLD and PV are combined together to handle with the music signal. In order to appropriately weighting the respective amplitude of NLD and PV, a transient content detector (TCD) is required to analyze successive time domain windows of the input signal. However, the TCD employed in this system has a poor performance of transient detection and the smearing of percussive still exists. Meanwhile, this system is very time costing. For instance, a 1-min music clip needs 2.38 mins [10] to accomplish the bass effect process on a PC. (Win 7, CPU: 3.4GHz, RAM: 4GB) Obviously, on portable devices which have limited computational abilities, the processing time as mentioned above is intolerable. Therefore, the real-time processing of bass effect is the most concerned in VBS.

In order to solve the problems described above, we propose a new hybrid VBS with high efficiency and better stead-state and percussive performance. This method is based on the concept of processing the percussive and the steady-state signal components separately with a harmonic and percussive sound separation (HPSS) [11] algorithm. We also employ an

improved PV based on the spectral peak detection to reduce the distortion in the generated harmonics and a direct FFT/IFFT based phase vocoder to decrease the processing time.



This paper is organized as follows. Section 2 describes the new VBS we proposed. Section 3 shows the subjective listening experiments results and the conclusion is presidented in Section 4.

Figure 1. Block diagram of the proposed virtual bass system.

II. Proposed Virtual Bass System

Figure 1 shows the block diagram of the proposed virtual bass system. L and R denote the left and right channel of the stereo input signals. Since the low-frequency signal generally lack of direction cues, these two signals filtered with a low-pass filter are combined into a mono signal with a 50% attenuation for low-frequency processing. The HPSS is to separate the steady-state and percussive signal components and then processing them through improved PV and NLD, respectively. For the low-frequency range of the percussive component contains limited signal energy that lead to reduced intermodulation distortion. And the high quality harmonics of the steady-state components are generated using an improved PV algorithms. After that, the processed signals go through a harmonic magnitude control (HMC) module G to ensure equal loudness. Finally, the two processed signal components are mixed and added to the high-pass original signal with delays. Then the output signals are the signals with bass effect.

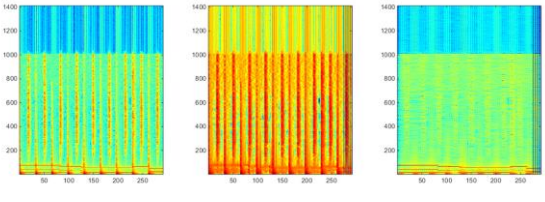


Figure 2. Steady-state and percussive signal components separaion using the HPSS. (a) The spectrogram of a music signal with steady-state and percussive components. The spectrogram of the separated percussive components (b) and the steady-state components (c).

A. Harmonic and Percussive Sound Separation

In our proposed system, we use the harmonic and percussive sound separation algorithms introduced in [11], which allows for real-time processing due to its low computational complexity. The basic concept of this method is that it focuses on the differences in the directions of continuity between the spectrograms of harmonic and percussive components. It means that the spectrograms of harmonic components are typically continuous in the time direction (as shown in Figure 2(c)), owing to their frequency stationary. The spectrograms of percussive components are typically continuous in the frequency direction, owing to their impulsiveness (as shown in Figure 2(b)). In [11], the source separation methods are formulated as optimization problems that optimize the "anisotropic smoothness" under some conditions.

Firstly, applying the short-time Fourier transform (STFT) to the original signal $y(t)$ and we get its spectrogram $\hat{\mathbf{Y}}$. In this algorithm, we just regard the spectrogram $\hat{\mathbf{Y}}$ as a tuple of $N \times K$ complex/real numbers, and define arithmetic operations as element-wise operations. So the amplitude spectrogram is $\mathbf{Y} = |\hat{\mathbf{Y}}| \in \mathbb{R}^{N \times K}$ (as shown in Figure 2(a)). Obviously, the spectrogram above is composed of $N \times K$ bins, where N is the number of time frames, and K is the number of bins in a single frame.

The "anisotropic smoothness" is describe as follows. Define $\mathbf{H} \in \mathbb{R}^{N \times K}$ as the amplitude spectrogram of steady-state signal components of $\mathbf{Y}$. We may assume that the value of the spectrogram at a time-frequency bin (n,k) ($n \in N, k \in K$), i.e. $H_{n,k}$, should be nearly equal to those of the temporally adjacent bins $(n \pm n', k)$. That is,

$$H_{n,k} \approx H_{n \pm n',k} \quad (1 \le n' \le N') \tag{1}$$

where N' is the maximal distance we consider neighbor from 1 to several dozen.

Similarly, the spectrogram of percussive signal components $\mathbf{P} \in \mathbb{R}^{N \times K}$ should have following property like (1),

$$P_{n,k} \approx P_{n,k \pm k'} \quad (1 \le k' \le K') \tag{2}$$

where K' is the maximal distance under consideration.

In order to measure the smoothness, we define the criteria as the sum of squared difference between the bins under consideration as follows,

$$S_{\text{time}}(\mathbf{H}^{\gamma}) = \sum_{n}\sum_{k}\frac{1}{N'}\sum_{n'=1}^{N'}(H_{n,k}^{\gamma} - H_{n-n',k}^{\gamma})^2 \tag{3}$$

$$S_{\text{freq}}(\mathbf{P}^{\gamma}) = \sum_{n}\sum_{k}\frac{1}{K'}\sum_{k'=1}^{K'}(P_{n,k}^{\gamma} - P_{n,k-k'}^{\gamma})^2 \tag{4}$$

Where γ is an exponential factor to suppress the effects from loud components and we define $\gamma = 0.5$ in this paper. Then, we define an integration of two criteria (3) and (4) as follows,

$$S(\mathbf{H}^{\gamma}, \mathbf{P}^{\gamma}; \omega) = S_{\text{time}}(\mathbf{H}^{\gamma}) + \omega S_{\text{freq}}(\mathbf{P}^{\gamma}) \tag{5}$$

where ω is a weighing constant. And we call Equation 5 as the smoothness function.

As we can see from (3) and (4), if given a bin $Y_{n,k}^{\gamma}$ and it's whether "steady-state predominant," "percussive predominant," or "silent" for each bin. In this case, it would be reasonable to classifying steady-state predominant bins into $\mathbf{H}^{\gamma}$ and

percussive predominant bins into $\mathbf{P}^{\gamma}$ results in smaller S. So the separation algorithm is actually the reverse of this

Problem :

minimize $S(\mathbf{H}^{\gamma}, \mathbf{P}^{\gamma}; \omega)$

subject to $\mathbf{H}^{2\gamma} + \mathbf{P}^{2\gamma} = \mathbf{Y}^{2\gamma}$

$\mathbf{H}^{\gamma} \geq 0, \mathbf{P}^{\gamma} \geq 0$

procedure, that is by minimizing the smoothness function $S(\mathbf{H}^{\gamma}, \mathbf{P}^{\gamma}; \omega)$, most of steady-state and percussive components may be classified into $\mathbf{H}^{\gamma}$ and $\mathbf{P}^{\gamma}$.

With the concept above, the source separation problem can be interpreted as an optimization problem as follows.

The constraint $\mathbf{H}^{\gamma} \geq 0, \mathbf{P}^{\gamma} \geq 0$ in the optimization problem indicate that the amplitude $H_{n,k}$ and $P_{n,k}$ cannot be negative. Another constraint $\mathbf{H}^{2\gamma} + \mathbf{P}^{2\gamma} = \mathbf{Y}^{2\gamma}$ depict the additivity of the amplitude spectrograms because of the wave domain additivity $h(t) + p(t) = y(t)$, where $h(t)$ and $p(t)$ are the steady-state and percussive signals, respectively.

As the optimization problems worked out, the steady-state and percussive signal components will be separated. Then, the steady-state signal components are processed through the improved PV and the percussive signal components through the NLD.

B. Improved Phase Vocoder

In previous virtual bass systems [8] [9], the conventional PV is used to generate the needed harmonics. In case the frequency range needs to be enhanced is $[f_{\text{low}}, f_{\text{cut}}]$, f_{low} is the lowest frequency to be boosted and f_{cut} is the cutoff frequency of the loudspeaker. For a local peak f_{p}, there exists an integer l which satisfies $(l-1)f_{\text{p}} < f_{\text{cut}}$ and $lf_{\text{p}} > f_{\text{cut}}$. Then we can generate the harmonic frequencies lf_{p}, $(l+1)f_{\text{p}}$, $(l+2)f_{\text{p}}$ of f_{p} (previous study [12] has shown that boosting the first 3 harmonic peaks above the cutoff frequency can achieve good bass effect), where l is the order of the lowest harmonic frequency above f_{cut}.

However, the phase consistency across different frequency bins cannot be guaranteed which could result in unnatural phaseness or reverberation and will influence the bass effect. In this paper, we directly boost the energy of original signal spectrogram at the harmonic frequency points lf_{p}, $(l+1)f_{\text{p}}$ and $(l+2)f_{\text{p}}$ so that the phase consistency can be certified. But another problem occurs, if the frequencies we boosted on the original spectrogram are valley points, boosting the energy of these points will destroy the peak point consistency of processed signal and original signal which can be easily perceived. In this case, a harmonic peak matching (HPM) algorithm is applied. For a specific harmonic frequency lf_{p} on the original spectrogram, we search for the nearest peak around it. As shown in Figure 3, if we find the peak at frequency $lf_{\text{p}} + \delta$ then we take the phase of $lf_{\text{p}} + \delta$ as the phase of lf_{p}. Same procedure is applied for higher harmonic.

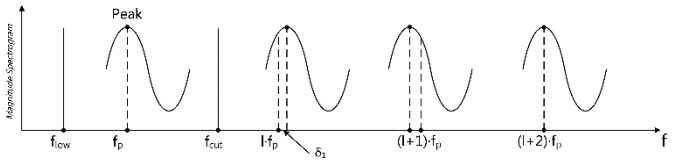


Figure 3. Illustration of harmonic peak matching.

C. Harmonic Magnitude Control

As the harmonic signals generated, the virtual bass system usually requires some mechanism to adjust the magnitude of each harmonic component to reflect proper timber and loudness. MaxxBass introduced a loudness analyzer to determine the weighting for each harmonic based on the SPL-to-phon expansion ratio $R(f)$ [13]. The SPL-to-phon expansion ratio is a function of the frequency f and can be represented as

$$R(f) = \frac{1.0}{\ln(f) \cdot 0.241 - 0.579} \tag{6}$$

The SPL-to-phon expansion ratio for the lth harmonic is given by

$$RR(f, l) = 1 - \ln(l) \cdot 0.241 \cdot R(f) \tag{7}$$

The energy of fundamental frequency and harmonic frequency meets the following equation [14]

$$E_h(f, l) = RR(f, l) \cdot E_f + k \tag{8}$$

Then we get

$$X_{lf}^2 = 10^{\frac{k}{10}} \cdot (X_f^2)^{RR(f,l)} \tag{9}$$

Where X_{lf} is the magnitude of lth harmonic frequency $l \cdot f$ and X_f is the magnitude of fundamental frequency f.

However, previous study [10] has shown that the added virtual bass will affect the perceived quality of the music. Specifically, the magnitude of harmonics calculated by Eq. (9) may affect the audio quality. Through subjective experiment, Zhou et. al. [15] find the relationship between the timber of bass enhanced signal and the relative intensity between harmonics. The result shows that when the relative intensity of different harmonics satisfies 0.5^x exponential decay (x is the difference of harmonic order), the perceived timber of bass enhanced music is the best. Based on this study, Zhang et. al. [16] proposed a subjective preference based HMC which can obtain good bass effect and maintain high audio quality at the same time. In his method, the lowest order harmonic magnitude X_{lf} is calculated first, and then calculate the other harmonic magnitude X_{mf} by

$$X_{mf} = X_{lf} \cdot 0.5^{(m-l)} \tag{10}$$

Because this method just needs to calculate the first harmonic magnitude using the approach in MaxxBass and other harmonic magnitude can be easily obtained by Eq. (10), the computation efficiency is subsequently improved.

III. EXPERIMENTS

The Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) subjective testing method [17], which aimed to meet the needs of researchers in evaluation audio systems that cannot be classified objectivity, is used for compare the performance of different bass enhancement systems. The bass enhancement system should be less sensitive to signal content, so we give consistent ratings across varieties of music genre. Seven 30-second music clips of different genre are used as the reference stimuli. The processed stimuli from NLD, PV and the proposed system are tested. The unprocessed signal is treated as the reference and the high-passed signal with 120 Hz cutoff frequency as the anchor. Seven subjects were asked to rate the stimuli from 0 (bad) to 100 (excellent). Test was carried out in a quite room where subjects were left alone to complete the test with no time constrains. Subjects listened to the stimuli over a Bose OE2 headphone driven by the REM FIREFACE UC USB high speed audio interface.

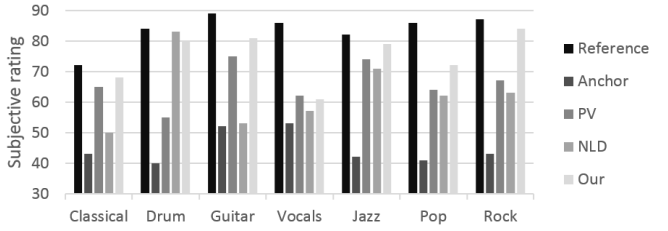


Figure 4. Average subjective ratings for bass effect.

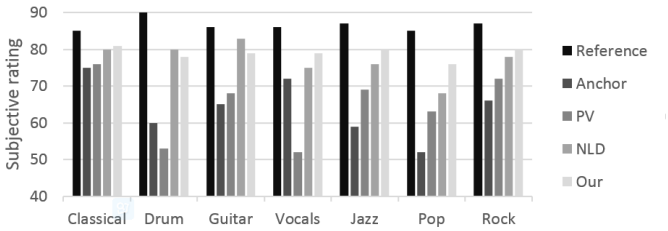


Figure 5. Average subjective ratings for audio quality.

The experiment results are shown in Figure 4 and Figure 5. As shown in Figure 4, the anchor has the lowest score for it has no low frequency components. Besides the reference, the proposed system has the highest score except the drum. A probable explanation is that drum signals have lots of percussive components which is more suitable for NLD processing. The audio quality is illustrated in Figure 5, from which we can see that the anchor has the lowest score too for the low frequency component is an essential of audio quality. However, the PV approach gets a relatively low score among different music genre. This is because the conventional PV has the phaseness or reverberation problem which would change the timbre and affect the audio quality, but the improved PV and subjective based HMC used in proposed system has solved this problem.

The computation time of different virtual bass approaches is shown in Table 1. The test is on a PC (Win 7, CPU: 3.4GHz, RAM: 4GB) and all approaches are coded using MATLAB. A 60-second music clip is taken as the test sample and the processing time is recorded. From the table we can see the approach we proposed has a very short processing time compared with other approaches, this makes it possible for real-time application.

TABLE I. COMPUTATION TIME OF DIFFERENT VBS.

Approach	PV	NLD	Hybrid	New-Hybird
Time (s)	96	57	156	36

IV. CONCLUSIONS

In this paper, we proposed a new hybrid virtual bass system based on the missing fundamental phenomenon, which utilizes a harmonic and percussive sound separation algorithm and subjective based HMC. From the subjective evaluation of the system we have done, the proposed VBS system produces more impactful bass effect with better audio quality. Moreover, this method improves the computation time dramatically with the help of the simplicity of the HPSS.

REFERENCES

[1] Fastl, Hugo, and Eberhard Zwicker. "Psychoacoustics: facts and models." (2001).

[2] N. Oo and W. S. Gan, "Harmonic and Intermodulation Analysis of Nonlinear Devices Used in Virtual Bass Systems," in *AES 124th Convention, Amsterdam, The Netherlands*, 2008.

[3] N. Oo, W. S. Gan, and W. T. Lim, "Generalized harmonic analysis of Arc-Tangent Square Root (ATSR) nonlinear device for virtual bass system," in 35th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2010, pp. 301–304.

[4] N. Oo and W. S. Gan, "Analytical and perceptual evaluation of nonlinear devices for virtual bass system," in 128th Convention of the Audio Engineering Society, London, UK, 2010.

[5] 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.

[6] Glotter, Daniel, and Meir Shashoua. "Method and system for enhancing quality of sound signal." U.S. Patent No. 5,930,373. 27 Jul. 1999.

[7]

[8] Bai, Mingsian R., and Wan-Chi Lin. "Synthesis and implementation of virtual bass system with a phase-vocoder approach." *Journal of the Audio Engineering Society* 54.11 (2006): 1077-1091.

[9] Hill, Adam J., and Malcolm OJ Hawksford. "A hybrid virtual bass system for optimized steady-state and transient performance." *Computer Science and Electronic Engineering Conference (CEEC), 2010 2nd*. IEEE, 2010.

[10] A. J. Hill, "Virtual bass toolbox," http://www.adamjhill.comlvb.html.

[11] Tachibana H, Ono N, Kameoka H, et al. Harmonic/Percussive Sound Separation based on Anisotropic Smoothness of Spectrograms[J]. 2014.

[12] Zhou and Z. Xie, 'The relationship between timbre of virtual bass and its components," Voice Technology, 2010.

[13] M. Shashoua and D. Glotter, "Method and System for Enhancing Quality of Sound Signal," US Patent 5930373 (1999).

[14] Ben-Tzur D, Colloms M. The effect of MaxxBass psychoacoustic bass enhancement on loudspeaker design[C]//Audio Engineering Society Convention 106. Audio Engineering Society, 1999.

[15] J. Zhou and Z. Xie, 'The relationship between timbre of virtual bass and its components," Voice Technology, 2010.

[16] Zhang S, Xie L, Fu Z H, et al. A hybrid virtual bass system with improved phase vocoder and high efficiency[C]//Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on. IEEE, 2014: 401-405.

[17] ITUR, "Bs. 1534-1. method for the subjective assessment of intermediate sound quality (mushra)," International Telecommunication Union, 2003.