Virtual Bass System Based on Rhythm Content Extracting

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. In this paper, we proposed a virtual bass system by extracting the rhythm content to improve the bass effect as well as maintain the audio quality. 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; 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 nolinear 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. According to this phenomenon the fundamental frequency can be reconstructed by the human central auditory system on the basis of higher order harmonics even if the fundamental partial is missing in the signal spectrum.

Compare with the standard bass boost techniques, the virtual bass system (VBS) have many advantages. Firstly, it enables to perceive low frequencies even if loudspeakers have no capability to reproduce them. Moreover, the effect of utilizing such an algorithm in small loudspeakers is generally better than using normal bass boost techniques. Finally, by using the VBS technique the risk of damaging the hardware can be avoided.

This paper proposed a VBS with extracting rhythm content. Section 2 shortly recalls the theoretical background of the missing fundamental phenomenon and the standard VBS. Section 3 makes a detailed description of the new VBS we proposed. Section 4 shows the subjective listening experiments results and the conclusion is presented in Section 5.

II. Theoretical Background

A. *Missing Fundamental*

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. For example if the spectral frequencies are at 100 Hz, 200 Hz, 300 Hz, 400 Hz etc, the perceived pitch will be 100 Hz for that signal. As a result, we are able to simulate the lower frequency pitches through their upper harmonics.

B. *Standard VBS*

On the basis of the missing fundamental phenomenon described above, the normal VBS utilizes the non-linear devices (NLD) or the phase vocoder (PV) to generate harmonics.

Non-linear devices [2-5] are the most commonly technique of harmonic generation used in VBS 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 intermodulation distortion (IMD), i.e. caused by two closely-spaced spectral components in the input signal. 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.

Another well-known method is phase vocoder [8]. It is operated in the frequency domain and allows for accurate control over the individual harmonic component. 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 switch between the NLD and PV a transient content detector (TCD) is employed. However, the TCD 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 example, a 1-min music clip needs 2.38 minutes [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 is intolerable. Therefore, the real-time processing of bass effect is an important concern in VBS.

Unlike those existing methods, the proposed approach greatly removes the disadvantages of both NLD and PV creating a better virtual bass system.

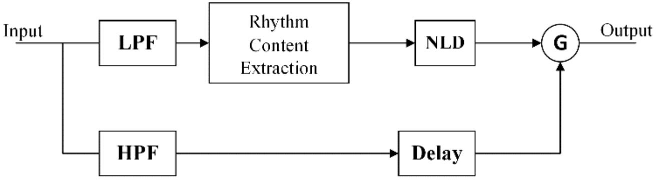

Figure 1. Proposed Virtual Bass System .

III. Proposed Virtual Bass System

Although NLD causes the IMD, the bass effect is more profound in NLD approach compared to PV. The real problem is how to generate good bass effect and reduce the IMD in the same time. As we know that NLD approach makes good performance for percussive sound such as drum beats. So we could get the drum beats components using the rhythm content extraction algorithm and processing them through NLD approach.

Figure 1 shows the block diagram of the proposed virtual bass system. Currently the system is implemented for mono channel, but it is easily scaled to dual channels.

The detailed description of each of the blocks in the system block diagram is as follows.

A. LPF and HPF

Since the virtual bass enhancement algorithm operate only the low frequency region. So the input signal needs to pass through a low pass filter (LPF). Because the cutoff has to be sharp, a 4th order elliptical FIR filter is chosen with a cutoff frequency of 120 Hz. The high frequency region of the input signal needs to be mixed with the processed low frequency portion. So the input signal has to pass through a high pass filter (HPF) and add some delays.

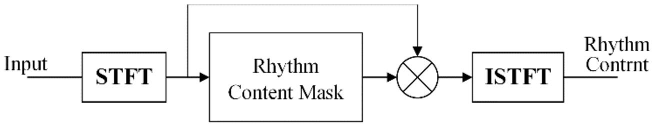

Figure 2. Rhythm Content Extraction.

B. Rhythm Content Extraction

To extract the rhythm content, the rhythm content extraction algorithm is used. Figure 2 display the procedure of extraction algorithm. Firstly, we applying the short-time Fourier transform (STFT) to the input signal. So we can get a Time-Frequency representation which assume the signal to be stationary for a narrow time frame. In this paper the signal is initially framed and windowed at 64 msec at a sampling frequency of 44.1 KHz. Overlapping ratio is 75%. The N = 1024 point FFT is applied on each time frame window. After we get the spectrogram of input signal, a rhythm content mask is calculated by audio source separation algorithm named harmonic and percussive sound separation (HPSS) algorithms introduced in [11], which allows for real-time processing due to its low computational complexity.

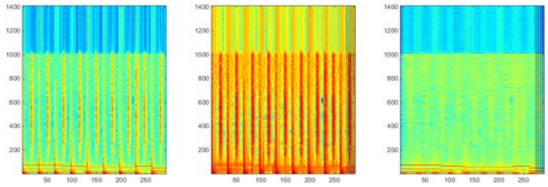

Figure 3. Steady-state and percussive signal components separaion using the HPSS. (a) The spectrogram of a music signal with steady-state and percussive components. The spectrogram of the separated percussive components (b) and the steady-state components (c).

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 3(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 3(b)). In [11], the source separation methods are formulated as optimization problems that optimize the "anisotropic smoothness" under some conditions.

Problem :

$$\begin{aligned} &\text{minimize} && S(\mathbf{H}^{\gamma}, \mathbf{P}^{\gamma}; \omega) \\ &\text{subject to} && \mathbf{H}^{2\gamma} + \mathbf{P}^{2\gamma} = \mathbf{Y}^{2\gamma} \\ &&& \mathbf{H}^{\gamma} \geq 0, \mathbf{P}^{\gamma} \geq 0 \end{aligned}$$

The block above is the optimization problem that can solve the audio source separation. The description of each symbol is as follows. $\mathbf{Y}$ is the spectrogram of input signal. $\mathbf{H}$ and $\mathbf{P}$ are the spectrogram of harmonic components and percussive components. γ is an exponential factor to suppress the effects from loud components and we define $\gamma = 0.5$ in this paper. Since the arithmetic in HPSS is element-wise operation so we regard the spectrogram just as a tuple of complex/real numbers. The constraint $\mathbf{H}^\gamma \geq 0, \mathbf{P}^\gamma \geq 0$ in the optimization problem indicate that the amplitude $\mathbf{H}$ and $\mathbf{P}$ cannot be negative. Because the input signal spectrogram is consist of harmonic and percussive portion, so the spectrogram is satisfy the additivity. Therefore, the constraint $\mathbf{H}^{2\gamma} + \mathbf{P}^{2\gamma} = \mathbf{Y}^{2\gamma}$ is obviously. $S(\mathbf{H}^\gamma, \mathbf{P}^\gamma; \omega)$ is the smoothness function of the optimization problem. It is defined as follows.

$$S(\mathbf{H}^\gamma, \mathbf{P}^\gamma; \omega) := S_{\text{time}}(\mathbf{H}^\gamma) + \omega S_{\text{freq}}(\mathbf{P}^\gamma) \tag{1}$$

where ω is a weighing constant. $S_{\text{time}}(\mathbf{H}^\gamma)$ and $S_{\text{freq}}(\mathbf{P}^\gamma)$ are defined as:

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

$$S_{\text{freq}}(\mathbf{P}^\gamma) := \sum_n \sum_k \frac{1}{K'} \sum_{k'=1}^{K'} (P^\gamma_{n,k} - P^\gamma_{n,k-k'})^2 \tag{3}$$

$H_{n,k}$ and $P_{n,k}$ are the value of the spectrogram at a Time-Frequency bin (n,k) ($n \in N, k \in K$). The equation (2) and (3) are criteria to measure the smoothness, it is defined as the sum of squared difference between the spectrogram bins $H_{n,k}$ and the adjacent bins $H_{n \pm n',k}$ under consideration. N' and K' are the maximal distance we consider neighbor from 1 to several dozen.

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 $\mathbf{P}^\gamma$. The rhythm content mask is generated using.

$$M(n,k) = \frac{P^\gamma(n,k)}{H^\gamma(n,k) + P^\gamma(n,k)} \tag{4}$$

Therefore, the rhythm content spectrogram can be extracted using

$$R(n,k) = Y(n,k) \cdot M(n,k) \tag{5}$$

Finally, we use inverse STFT to reverse the rhythm content spectrogram back to time domain signal.

C. Non-Linear Device (NLD)

With extracting the rhythm content the algorithm goes back to time domain operations. From a study of analysis of different nonlinear devices [4]. Arc Tangent Square Root (ATSR) performs well in good bass effect, so in our system we use ATSR as harmonic generator.

D. 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

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

Where $X_{l\,f}$ 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 $X_{l\,f}$ is calculated first, and then calculate the other harmonic magnitude $X_{m\,f}$ by

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

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.

IV. 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 quiet 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.

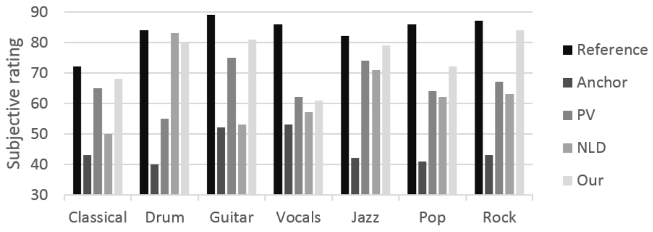

Figure 4. Average subjective ratings for bass effect.

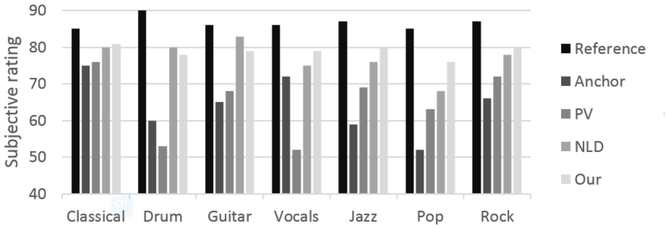

Figure 5. Average subjective ratings for audio quality.

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	Our system
Time (s)	96	57	156	36

V. CONCLUSIONS

In this paper, we proposed a VBS with extracting rhythm content based on the missing fundamental phenomenon. It utilizes a rhythm content extraction 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.