A VIRTUAL BASS SYSTEM WITH IMPROVED OVERFLOW CONTROL

Hao Mu and Woon-Seng Gan

Digital Signal Processing Lab, School of Electrical and Electronic Engineering,
Nanyang Technological University, Singapore

ABSTRACT

The virtual bass system (VBS) can enhance the bass performance of small or flat loudspeakers by tricking the human brain to perceive the fundamental frequency from its higher harmonics. However, additional harmonics may lead to arithmetic overflow and cause distortion due to clipping, especially during high-level transient components. Past research pay little attention on this problem, and manual control of VBS gain settings is required to prevent overflow. Users need to manually adjust the gain settings for different sound tracks, which can be very troublesome. In this paper, we propose a VBS that can efficiently prevent the overflow problem by automatically controlling the gain settings for additional harmonics. This new approach pre-computes the gain limitation for additional harmonics and can be adopted for real-time audio implementation. Objective measurements are carried out to compare the proposed method with the commonly used limiter method.

Index Terms— music signal processing, overflow control, psychoacoustic bass enhancement, transient detection

1. INTRODUCTION

The virtual bass system (VBS) [1]–[8] is a psychoacoustic method to enhance the bass performance of small or flat loudspeakers that are not capable of reproducing good bass (low-frequency) effect due to their high cut-off frequency. The VBS tricks human brains to perceive the low-frequency component that is not physically reproduced. It is based on a psychoacoustic phenomenon known as the "missing fundamental" [9], [10], which states that higher harmonics of the fundamental frequency can reproduce the sensation of the fundamental frequency in the human auditory system. Therefore, VBS can virtually reproduce the sensation of f_0 in a loudspeaker having a cut-off frequency higher than f_0 by injecting a series of harmonics at frequencies of $2f_0$, $3f_0 \ldots 6f_0$, as shown in Figure 1. A high-passed filter (HPF in Figure 1) is applied to the original signal to remove the redundant low-frequency components that cannot be reproduced by the loudspeaker.

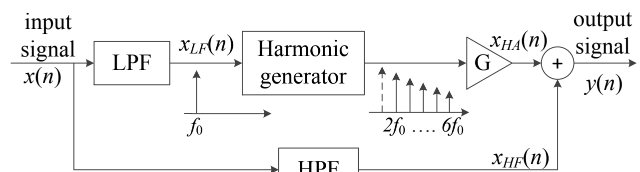

Figure 1. General framework of the VBS. (LPF and HPF: low-pass and high-pass filters; G: gain for harmonics; $x(n)$ and $y(n)$: input and out signal, where n represents the sample index; $x_{LF}(n)$ and $x_{HF}(n)$: low-frequency and high-frequency components of the input signal; f_0: the fundamental frequency; $x_{HA}(n)$: synthesized harmonics).

Our previous study proposed a hybrid VBS [6] that separates input signal into steady-state and transient components and a timbre matching approach [8] to improve the audio quality of the VBS-enhanced signals. However, due to the arithmetic addition of the high-pass filtered signals $x_{HF}(n)$ and additional harmonics $x_{HA}(n)$, there is a possibility of arithmetic overflow at the output signal $y(n)$, especially for high-amplitude transient signals. This results in clipping distortion in the playback (as shown in Figure 2), which is usually perceived as harsh sounds [11].

A common method to prevent overflow is to use a limiter after VBS. The limiter is a type of dynamic range compressor (DRC) [12], [13]. It provides attenuation over the highest amplitude level in the signal and changes the dynamics of low-level signals as little as possible. The attenuation gain is computed by comparing the detected level of $y(n)$ and the threshold of the limiter. However, the limiter also attenuates the high frequency components of $x_{HF}(n)$ that can be physically reproduced, and degrades the output signal. Furthermore, it is difficult for users to choose parameters for the limiter to ensure both compression effect and low distortion.

Instead of controlling the output levels, Larsen and Arts [14] introduced a feedback method of controlling the gain for additional harmonics in response to the level of the output signal. Different from the limiter, this method does not affect $x_{HF}(n)$. However, details of this feedback controlling method were not proposed, and its performance was not evaluated. Extension from their idea, we propose an automatic gain control method to prevent overflow in VBS. As shown in Figure 3, the proposed VBS with gain control is extended from our previously proposed hybrid VBS [6].

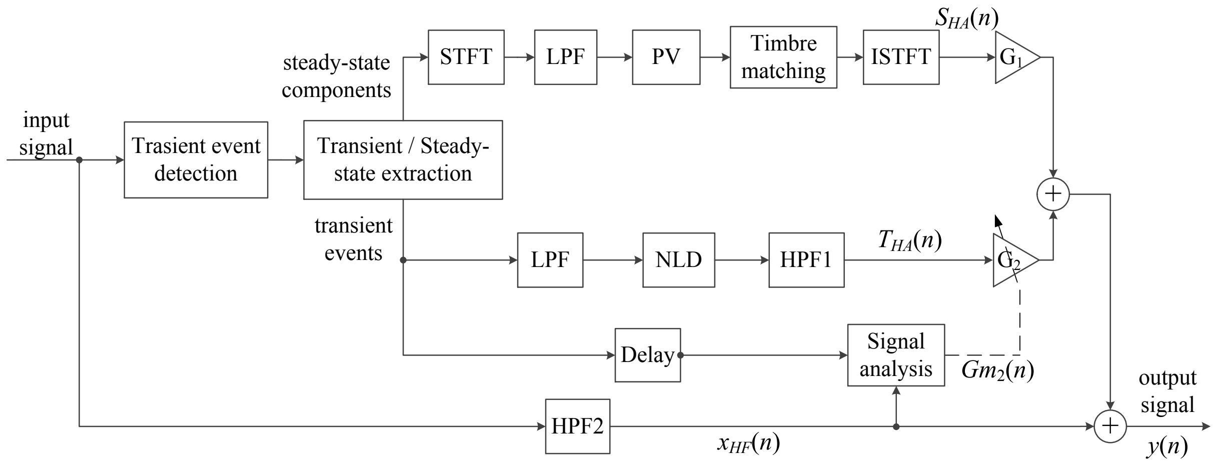

Figure 3. General framework of the proposed VBS with gain control. (STFT and ISTFT: short-time Fourier transform and inverse short-time Fourier transform; LPF and HPF: low-pass and high-pass filters; PV and NLD: the phase vocoder and the nonlinear device; $x_{HF}(n)$: high-frequency components of the input signal; $S_{HA}(n)$ and $T_{HA}(n)$: synthesized harmonics of steady-state and transient components; $Gm_2(n)$: maximum gain for harmonics).

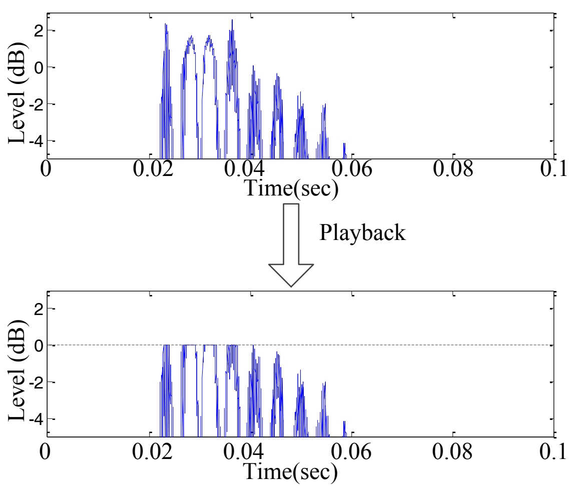

Figure 2. Clipping distortion in the playback due to the arithmetic overflow of the signal.

Since overflow is much more likely to occur in high level signals, the gain control is only applied in transient components. A transient event detection method is proposed based on our previous separation method for steady-state and transient components [6]. The maximum gain $Gm_2(n)$ for harmonics in each transient event are calculated by analyzing the levels of high-pass filtered signals $x_{HF}(n)$ and transient additional harmonics $T_{HA}(n)$.

The remainder of this paper is organized as follows. In Section 2, we introduce the transient event detection method. Section 3 presents the gain control method, and Section 4 compares the performance of the proposed gain control method and the limiter. The paper is concluded in Section 5.

2. DETECTION OF TRANSIENT EVENTS

The proposed gain control method is based on the detection of transient events. In our previous research [6], a separation method of steady-state and transient signals based on median-filters [15] was used. Spectrogram masks are produced to divide spectrogram elements into the two groups. However, the separated transient signal usually contain some residual steady-state components (as shown in Figure 4), especially during the attacks of steady-state components [16], which influences the following gain controlling method.

Instead of separating spectrogram elements, the proposed VBS separates steady-state and transient components by segmenting the frames of the input signal. Processing blocks of the proposed separation method is shown in Figure 5. It consists of a cascade system based on the median-filter-based separation method [15], follows by the transient event detection using the high frequency content (HFC) function. The HFC function is defined as:

$$HFC(m) = \frac{1}{K}\sum_{k=1}^{K} k \cdot \left|X(m,k)\right|^2, \qquad (1)$$

where m and k represent time frame and frequency bin indices, respectively; $X(m,k)$ denotes the spectral component, and K is the spectrum length. The HFC function has been successfully used in the detection of percussive onsets [17], where transient signals are modeled as a vertical ridge in the spectrogram.

As shown in Figure 6, there are 3 steps for the detection of transient events based on the HFC function: (i) Detect the peak frame m_{peak} of HFC, which indicates peak levels of transient events. (ii) Detect the onset frame m_{onset} by finding the HFC notch in 15 frames before m_{peak}. (iii) Detect the offset frame m_{offset} by finding the HFC notch in 30 frames (197ms) after m_{onset}. In [18], FitzGerald *et al.* suggested to use the minimum and maximum lengths of 50ms and 200ms in segmenting percussive (or transient-like) signals, which guarantees enough information for the following feature extraction. Therefore, our selected period of 30 frames is sufficient to capture transient events. In addition, half length

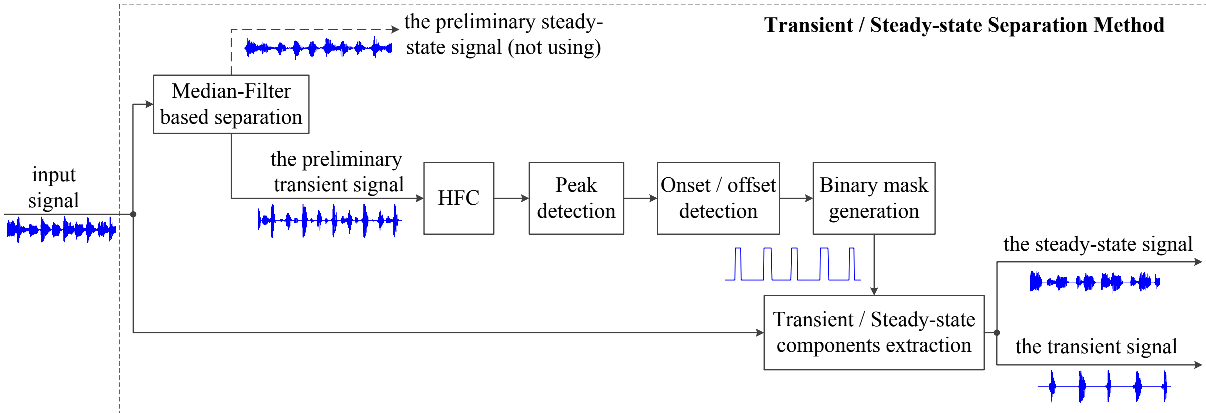

Figure 5. Processing blocks of the proposed steady-state and transient separation method.

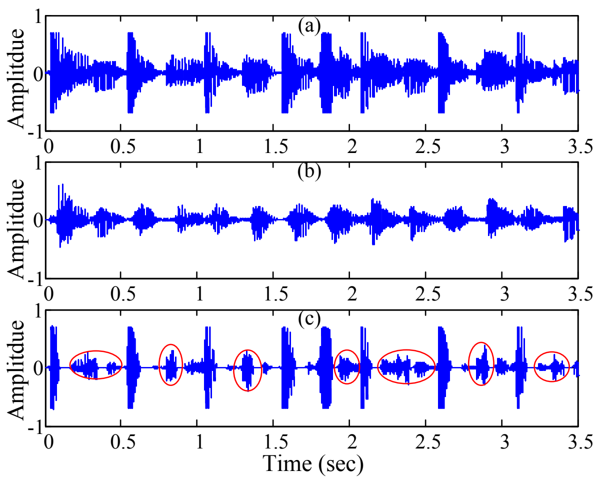

Figure 4. Steady-state and transient separation using median-filter based method. a) The input signal. b) The separated steady-state signal. c) The separated transient signal. (Note: the circles indicate steady-state components contained in the separated transient signal.

(15 frames) of the transient event is sufficient to detect the onset frame, as the attack time of a transient event is generally faster than the release time.

Subsequently, the detected onset and offset frames are used to generate temporal binary masks for input signals (as shown in Figure 5). The indices of onset and offset frames (m_{onset} and m_{offset}) are transformed into sample indices (n_{onset} and n_{offset}) using the inverse short-time Fourier transform (ISTFT):

$$\begin{aligned} n_{onset} &= m_{onset} \cdot n_{hop} + 1, \\ n_{offset} &= m_{offset} \cdot n_{hop} + n_{win}, \end{aligned} \tag{2}$$

where n_{hop} and n_{win} are hop size and window size used in the short-time Fourier transform (STFT). Input samples between pairs of n_{hop} and n_{win} are identified as transient events (masked as 1), while the others are steady-stated components (masked as 0).

3. GAIN CONTROL METHOD

After detection of transient events, the maximum gain $Gm_2(n)$ for harmonics is calculated for each transient event. By setting the gain for harmonics below $Gm_2(n)$, levels of the output signal $y(n)$ is controlled below the digital full scale (0 dB). When users set the gain for harmonics larger than $Gm_2(n)$ of a transient event, the gain will be reduced and fixed to $Gm_2(n)$ during the transient event.

The maximum gain for harmonics $Gm_2(n)$ is derived from the levels of transient additional harmonics $T_{HA}(n)$ and high-pass-filtered signal $x_{HF}(n)$. In one transient event from n_{onset} to n_{offset}, $Gm_2(n)$ should ensure that the arithmetic addition of $T_{HA}(n)$ and $x_{HF}(n)$ does not exceed the full scale level 0 dB:

$$\max_i \left\{ Gm_2(i) \cdot |T_{HA}(i)| + |x_{HF}(i)| \right\} = 0 \text{ dB}, \tag{3}$$

where $i = n_{onset} \ldots n_{offset}$. From (3), we can derive the maximum gain for harmonics:

$$Gm_2(i) = \min_i \left\{ \frac{0 \text{ dB} - |x_{HF}(i)|}{|T_{HA}(i)|} \right\}, \tag{4}$$

where $i = n_{onset} \ldots n_{offset}$.

In our informal tests, it is found that the peak of the transient event is the most possible place to have overflow distortion, so it is possible that $Gm_2(n)$ derived from part of the transient events is sufficient to cover the whole event. The range of i in (4) can be changed to $i = n_{onset} \ldots n_{end}$, where n_{end} determines how many samples of the whole transient event is used to determine $Gm_2(n)$. We can use part of the transient event ($n_{end} < n_{offset}$), or the whole event ($n_{end} = n_{offset}$). The computation of $Gm_2(n)$ uses the samples from n_{onset} to n_{end}, which results in a delay time in the proposed method:

$$\text{Delay time} = (n_{end} - n_{onset}) \cdot T_s \text{(sec)}, \tag{5}$$

where T_s is the sampling period. When $n_{end} = n_{offset}$, the maximum delay time equals to the detection range of the transient event (197ms). The minimum delay time equals to the detection range of the onset frame (15 frames or 110ms). The performances of different delay time settings will be discussed in the next section.

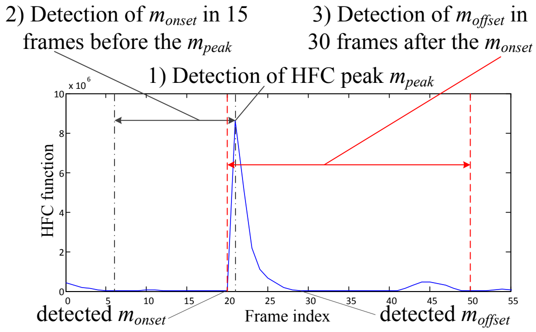

Figure 6. Detection of transient events using the HFC function. (m_{peak}: the peak frame, m_{onset}: the onset frame, m_{offset}: the offset frame; − · −: detection range for m_{onset}, − − −: detection range for m_{offset})

4. COMPARING THE GAIN CONTROL METHOD WITH THE LIMITER

As mentioned in Section 1, the limiter can also be used to prevent overflow in VBS. Compared to our proposed method, the limiter does not result in delay to the system. However, the limiter requires users to choose the threshold value, which may lead to different results for different sound tracks. In this section, we evaluate the overflow control performances of the proposed gain control method with different delays and the limiter with different thresholds. Twenty polyphonic stimuli with sufficient low-frequency components from the music audio benchmark data set [19] and 5 single instrument (bass drum solo) stimuli are selected for our experiment. Additional harmonics are generated during the detected transient events, and the gain for harmonics is set to 6 dB. In total, 239 transient events are detected in the 25 stimuli, and out of which, 225 transient events have overflow.

As mentioned in Section 4, the delay time of the proposed gain control method is determined by the selected n_{end}. We tested 5 different delay time settings, and the result is listed in Table 1. The performances of all the delay settings are acceptable. The shorter delay time leads to more overflow. No overflow occurred using the delay time above 174ms, and the minimum delay time of 110ms resulted in 5.33% overflow transient events. Users can choose the delay time according to their applications. For audio-only applications, longer delay time can be used to completely prevent the overflow in VBS, and the delay of 174ms is sufficient for real-time implementation. In the audio/video playback, the delay may result in audio/video sync error. ITU-R BT.1359-1 [20] proposes a detectability threshold of 125ms and an acceptability threshold of 185ms for audio/video sync error. Hence, the delay of 122ms can be used for undetectable sync error with very a small possibility of overflow (3.11%), or the delay of 174ms can be used to completely prevent overflow with acceptable audio/video sync error.

Table 1. List of overflow testing results of the proposed gain control method with different delay time. (The percentage of overflow is computed by dividing the resulted number of overflow transient events by the total 225 overflow transient events).

Setting delay	Number of over flow transient events	Percentage of overflow
197ms	0	0%
174ms	0	0%
151ms	2	0.89%
122ms	7	3.11%
110ms	12	5.33%

Table 2. List of overflow testing results of the limiter with different thresholds. (The percentage of overflow is computed as same as Table 1).

Setting threshold	Numbers of overflow (percentage)	Average level of non-overflow
0 dB	224 (99.56%)	-0.30 dB
-3 dB	49 (21.87%)	-0.97 dB
-6 dB	0 (0%)	-1.78dB

The limiter function used in our test is obtained from the intelligent DRC MATLAB toolbox [12]. The attack time and release time are set to 1ms and 5ms, respectively. Three thresholds for the limiter, 0dB, -3dB and -6dB are tested, and the result is listed in Table 2. As the limiter generally requires a time to act, setting the threshold to 0 dB can hardly prevent overflow. It resulted in 99.56% overflow transient events in our test. Only the limiter with threshold of -6 dB can effectively prevent the overflow. However, a low threshold may overly attenuate the transient signal and result in a weaker perception of bass enhancement. Using the limiter with the threshold of -6 dB, the average level of the output signal was attenuated to -1.78 dB, which heavily reduced not only additional harmonics but also high-frequency components of the input signal, leading to degradation of the output signal.

5. CONCLUSION

In this paper, we proposed a harmonic gain control method to prevent overflow and clipping distortion that usually occur in VBS. A detection method of transient events is included to compute a suitable gain limitation for additional harmonics. The evaluation test results indicate that the proposed method can effectively prevent overflow with a small delay, which allows the system to be implemented in real-time applications. Compared to the common limiter method, the proposed method does not require users to adjust the parameters for different types of sound tracks. Our test revealed that the proposed gain control method can effectively prevent overflow and keeps the maximum level at the full scale (0 dB). The limiter only with low setting threshold can effectively prevent overflow, but the levels of output signals were overly attenuated.

6. REFERENCES

[1] W. S. Gan, S. M. Kuo, and C. W. Toh, "Virtual bass for home entertainment, multimedia PC, game station and portable audio systems," *IEEE Transactions on Consumer Electronics*, vol. 47, no. 4, pp. 787–796, 2001.

[2] N. Oo and W. S. Gan, "Harmonic and intermodulation analysis of nonlinear devices used in virtual bass systems," presented at the 124th Audio Eng. Soc. Conv., Amsterdam, 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] H. Mu, W. S. Gan, and E. L. Tan, "A psychoacoustic bass enhancement system with improved transient and steady-state performance," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 2012, pp. 141 –144.

[7] A. J. Hill and M. O. J. Hawksford, "A hybrid virtual bass system for optimized steady-state and transient performance," in *Proc. 2nd Computer Sci. and Electronic Eng. Conf. (CEEC)*, Colchester, UK, 2010, pp. 1,6, 8–9.

[8] H. Mu, W. S. Gan, and E. L. Tan, "A timbre matching approach to enhance audio quality of psychoacoustic bass enhancement system," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, 2013, pp. 36–40.

[9] J. F. Schouten, R. J. Ritsma, and B. L. Cardozo, "Pitch of the residue," *Journal of the Audio Engineering Society*, vol. 34, no. 9B, pp. 1418–1424, 1962.

[10] E. Terhardt, "Calculating virtual pitch," *Hearing Research*, vol. 1, no. 2, pp. 155–182, Mar. 1979.

[11] B. Katz and R. A. Katz, *Mastering audio: the art and the science*. Butterworth-Heinemann, 2003.

[12] D. Giannoulis, M. Massberg, and J. D. Reiss, "Digital dynamic range compressor design—A tutorial and analysis," *Journal of the Audio Engineering Society*, vol. 60, no. 6, pp. 399–408, 2012.

[13] U. Zölzer and X. Amatriain, *DAFX: digital audio effects*. John Wiley and Sons, 2002.

[14] E. R. Larsen and R. M. Aarts, *Audio bandwidth extension: application of psychoacoustics, signal processing and loudspeaker design*. Wiley, 2004.

[15] D. Fitzgerald, "Harmonic/Percussive Separation using Median Filtering," in *13th International Conference on Digital Audio Effects (DAFX10)*, Graz, Austria, 2010.

[16] H. Tachibana, N. Ono, H. Kameoka, and S. Sagayama, "Harmonic/Percussive Sound Separation based on Anisotropic Smoothness of Spectrograms," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[17] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, Sep. 2005.

[18] A. Klapuri and M. Davy, *Signal processing methods for music transcription*. Springer, 2006.

[19] H. Homburg, I. Mierswa, B. Möller, K. Morik, and M. Wurst, "A Benchmark Dataset for Audio Classification and Clustering.," in *ISMIR*, 2005, vol. 2005, pp. 528–31.

[20] "Relative Timing of Sound and Vision for Broadcast Operation," ITU-R Recommendation BT.1359-1, 1998.