A PSYCHOACOUSTIC BASS ENHANCEMENT SYSTEM WITH IMPROVED TRANSIENT AND STEADY-STATE PERFORMANCE

Hao Mu, Woon-Seng Gan, Ee-Leng Tan

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

ABSTRACT

Psychoacoustic bass (low frequency) enhancement approach has received strong interests from consumer electronics manufacturers who demand good bass quality from small or flat panel speakers. Due to physical limitations and cost constraints, good bass performance is lacking from such speakers. A typical solution is based on the psychoacoustic phenomenon known as the "missing fundamental", whereby human auditory system can perceive the fundamental frequency from its higher harmonics. This psychoacoustic bass enhancement system generally uses nonlinear devices (NLD) or phase vocoders (PV) to generate harmonics that enhance bass virtually. However, both approaches have their strengths and weaknesses. This paper presents a hybrid system, which combines these two approaches, to overcome their drawbacks and achieve good bass performance. The new approach first separates musical signals into transient and steady-state components, and applies NLD and PV on the separated signals. MUSHRA subjective test is used to evaluate the bass effect and audio quality of the proposed hybrid system against the NLD and PV methods.

Index Terms— virtual bass, psychoacoustics, NLD, phase vocoder, median filter

1. INTRODUCTION

The problem of the small-sized loudspeakers in consumer electronics devices is that they cannot produce good bass (low frequency) effect due to their physical size limitation and cost constraint. Gan *et al.* [1] introduced a bass enhancement system based on the psychoacoustic phenomenon called "missing fundamental" to enhance the bass effect of desktop speakers. The missing fundamental suggests that the higher harmonics of the fundamental frequency can produce the presence of the fundamental frequency in the human auditory system. The nonlinear device (NLD) and phase vocoder (PV) are commonly used to generate the higher harmonics. The generated harmonics are then recombined with the original signal so as to strengthen its bass perception psychoacoustically, as shown in Figure 1. This system is commonly known as the virtual bass system (VBS).

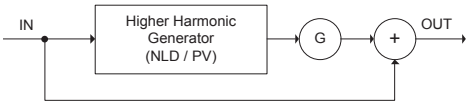

Figure 1. Virtual bass system.

A polynomial expansion of a particular function is generally used as a NLD in VBS. The polynomial approximated NLD can be represented by

$$y(n) = \sum_{i=0}^{N} h_i x^i(n), \tag{1}$$

where $y(n)$ is the output, $x(n)$ is the input, and h_i are the polynomial coefficients. From the previous study, NLD based VBS results in an impressive performance on the transient components such as the drum beats [2]. However, different NLDs suffer from different degrees of inter-modulation distortions, and an in-depth subjective measurement [3] must be carried out to assess audible distortion of NLD.

A PV based VBS has been introduced in [4]. This system operates in the frequency domain, and features precise control over individual harmonics. Therefore, the intermodulation problem is avoided, and the generated harmonics yields better audible quality compared to NLDs. The PV estimates the instantaneous frequency of the input signal by analyzing phase propagation of the spectrogram. The higher harmonics are reproduced as the sum of the sinusoids

$$y(n) = \sum_i \sum_k M(n,k) \cos(2\pi \cdot \alpha_i \cdot \hat{\omega}(n,k)), \tag{2}$$

where $M(n,k)$ is the analysis magnitude in time frame n and frequency channel k. $\hat{\omega}(n,k)$ is the estimated angular frequency, and α_i is the shifting parameter, which is also the generated harmonic's order. However, the PV based VBS generates transient smearing, which is perceived as a loss of percussiveness. Since the smearing of transient may cause unnatural effects in musical signals, PV is more applicable to steady-state signals than NLD.

To exploit the advantages of both NLD and PV, Hill *et al.* [2] designed a hybrid virtual bass system. This system requires a transient content detector to handle the mixing of the NLD's and PV's outputs. However, varying weights between transient and steady-state components may cause unnatural occasional switching effect and distort the quality of the sound track. Therefore, we propose a new hybrid virtual bass system, which overcomes the drawbacks of the prior approaches and achieves a better bass performance.

This paper is organized as follows. Section 2 describes our proposed hybrid virtual bass system. It highlights several key blocks that lead to the enhanced bass perception. Section 3 shows the result of the subjective test, and conclusion is presented in Section 4.

2. PROPOSED HYBRID VIRTUAL BASS SYSTEM

Figure 2 shows the block diagram of our proposed hybrid virtual bass system. Our hybrid system separates the signal spectrum into transient and steady-state components, and then applies NLD and PV on the respective signal components. We use a simple and yet effective median filter based separation method, as described in [5]. The separation masks for different signal components are generated from the median filters' outputs and the input spectrum are separated by respectively multiplying these two masks.

The separated transient signal goes through the NLD to generate higher harmonics for the transient component. In musical signal, low-frequency range of the transient component contains limited signal energy that leads to reduced inter-modulation distortion. On the other hand, high quality harmonics of the steady-state components are generated using an improved PV algorithm. Comparing with the conventional sum-of-sinusoids method utilized in [2] and [4], the improved PV is superior in phase handling.

2.1 Transient and steady-state signal separation

In the proposed system, we use the median filer based separation method introduced in [5], which permits real-time processing due to its low computational complexity. This method is an alternative to Ono's algorithm [6], which is based on the fact that the steady-state component appears as a horizontal ridge on the magnitude spectrogram, while the transient component forms a vertical ridge (Figure 3(a)). Therefore, when median filter is applied across the time axis, it smoothens out the horizontal (harmonic) lines and filters out the vertical (transient) lines, and produces the steady-state-enhanced components. Similarly, a transient-enhanced component is produced when median filter is applied across the frequency axis.

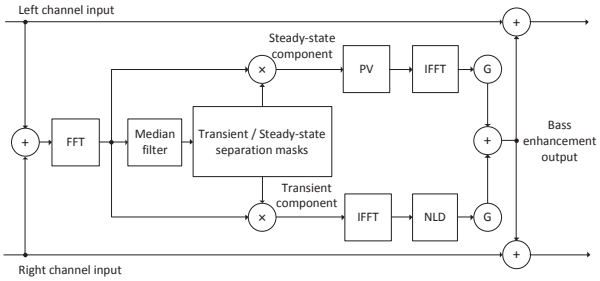

Figure 2. The block diagram of the proposed virtual bass system.

To avoid artifacts introduced by the nonlinearity of the median filter, the enhanced spectrum derived from the median filter is used to generate the soft separation masks. The transient-enhanced spectrogram is denoted by $T(n,k)$ and the steady-state-enhanced spectrogram is denoted by $S(n,k)$. The spectrum separation masks for the transient components $M_T(n,k)$ and the steady-state components $M_S(n,k)$ are generated using

$$M_T(n,k) = \frac{T^2(n,k)}{T^2(n,k) + S^2(n,k)}, \tag{3}$$

and

$$M_S(n,k) = \frac{S^2(n,k)}{T^2(n,k) + S^2(n,k)}. \tag{4}$$

Therefore, the transient and steady-state spectrograms can be respectively extracted using

$$\hat{T}(n,k) = X(n,k) \cdot M_T(n,k), \tag{5}$$

and

$$\hat{S}(n,k) = X(n,k) \cdot M_S(n,k), \tag{6}$$

where $X(n,k)$ denotes the input spectrograms. Figure 3(b) and (c) show the results of the proposed separation method on the spectrogram in Figure 3(a). We can see that the transient and steady-state spectrograms are clearly separated. Furthermore, since the masks only operate on the magnitude spectrum, the phase relationship between the separated spectrums is preserved.

2.2 Improved Phase Vocoder

The conventional PV used in [2] and [4] is based on an impractical assumption that every frequency channel corresponds to a constant-frequency sinusoid. This assumption leads to distortion when the input consists of frequency-sweeping signals (e.g. glissando and vibrato), as shown in Figure 4(b). Therefore, an improved PV based on the spectral peak detection is applied in the proposed system, which can significantly reduce the distortion in the generated harmonics of the musical signal.

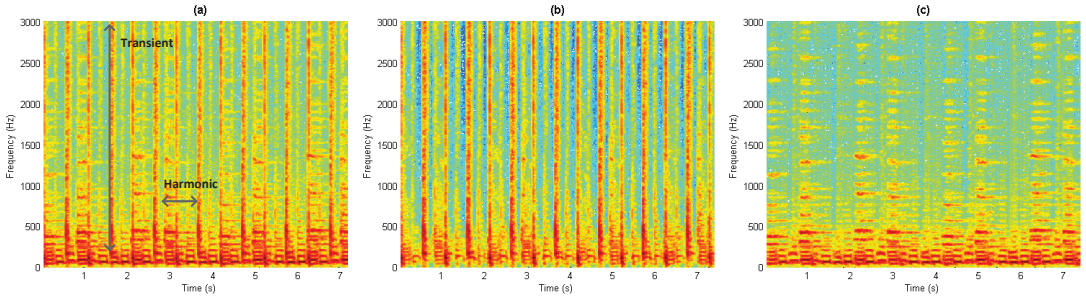

Figure 3. Transient and steady-state separation using the median filter approach. (a) The spectrum of a musical signal with transient and steady-state components. The spectrum of the separated (b) transient and (c) steady-state components.

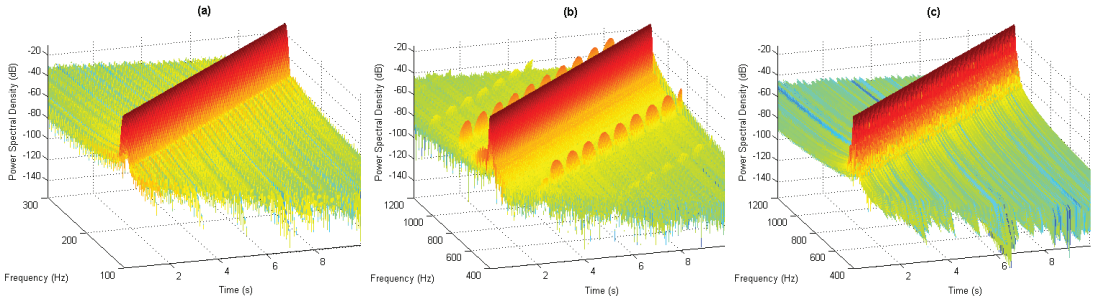

Figure 4. Harmonic generation using PVs. (a) spectrogram of the input signal, which is a linear chirp. (b) the spectrum of the 4th harmonic generated from the conventional PV. (c) the spectrum of the 4th harmonic generated from the improved PV.

According to the convolution theorem, the magnitude spectrum of a windowed sinusoid at frequency f_k is actually the spectrum of the analysis window with its main-lobe centered at frequency f_k, which is a peak on the magnitude spectrum [6]. Therefore, sinusoids in the signal can be represented by spectral peaks. In the proposed system, the spectral peaks are identified by picking the local maximum within a frequency range equal to the main-lobe bandwidth of the analysis window. Next, the spectrum is divided into influence regions centered around each peak. The border of the adjacent regions is set as the nulls between the two peaks. All frequency bins in the influence region contribute to the same sinusoid related to the peak. Following the idea in [7], the proposed PV generates the spectrum of the higher harmonics by shifting positions of the influence regions. For example, assuming that the instantaneous frequency of the spectral peak is $\hat{\omega}_p$, all the frequency bins in the corresponding region are shifted by the distance $(\alpha-1)\hat{\omega}_p$, where α is the harmonic's order. The reassignment method is used for robust estimation of the instantaneous frequency on the complex spectrum. Detailed information of the reassignment method is covered in [8].

Phase coherence also influences the sound quality of PV. Since spectral peaks represent the sinusoids of the signal, the steady-state signal $x(t)$ can be modeled as the sum of sinusoids

$$x(n) = \sum_{i=1}^{I(t)} A_i(n)\cos(\phi_i(n)),$$
$$\phi_i(n) = \phi_i(0) + \sum_{\tau=0}^{n} \tau \frac{\hat{\omega}_i(\tau)}{f_s}, \tag{7}$$

where $\phi_i(n)$, $A_i(n)$ and $\hat{\omega}_i(n)$ are the instantaneous phase, amplitude and frequency of the ith spectral peak, respectively. Assuming that the ith sinusoid is shifted from $\hat{\omega}_i(\tau)$ to $\alpha\cdot\hat{\omega}_i(\tau)$, we set the initial synthesis instantaneous phase $\phi_{si}(0) = \alpha\cdot\phi_i(0)$. The synthesis instantaneous phase of the shifted sinusoid becomes

$$\begin{aligned}\phi_{si}(n) &= \phi_{si}(0) + \sum_{\tau=0}^{n} \tau \frac{\alpha\cdot\hat{\omega}_i(\tau)}{f_s} \\ &= \alpha\cdot\phi_i(0) + \alpha\cdot\sum_{\tau=0}^{n} \tau \frac{\hat{\omega}_i(\tau)}{f_s} \\ &= \alpha\cdot[\phi_i(0) + \sum_{\tau=0}^{n} \tau \frac{\hat{\omega}_i(\tau)}{f_s}] \\ &= \alpha\cdot\phi_i(n).\end{aligned} \tag{8}$$

However, from the spectral peaks, we can only get the wrapped phase (restricting to $(-\pi,\pi]$). Assuming that the spectral peak k_l has the instantaneous phase $\phi(n,k_l)$, its wrapped phase is

$$\phi_{wrap}(n,k_l) = \phi(n,k_l) - 2m\pi, \tag{9}$$

where m is an integer. With the integer shifting parameter α, the synthesis phase becomes

$$\begin{aligned}\phi_s(n,k_l) &= \alpha\cdot\phi(n,k_l) \\ &= \alpha\cdot\phi_{wrap}(n,k_l) + 2\alpha m\pi.\end{aligned} \tag{10}$$

Since both α and m are integers, $2\alpha m\pi$ in (10) can be dropped. Therefore, the synthesis phase coherence in the proposed PV can be easily preserved by setting

$$\phi_s(n,k_l) = \alpha \cdot \phi_{wrap}(n,k_l). \quad (11)$$

The synthesis phase of the remaining frequency bins in each influence region are locked by

$$\phi_s(n,k_l) - \phi_s(n,k) = \phi_{wrap}(n,k_l) - \phi_{wrap}(n,k), \quad (12)$$

in order to reduce the "Phasiness" artifacts described in [9].

Consequently, the proposed phase vocoder shows a significant improvement on harmonics' quality. As shown in Figure 4(c), the distortion from the improved PV is significantly reduced as compared with the conventional PV.

3. SUBJECTIVE LISTENING TESTS

The Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) method was used for the subjective evaluation to compare the performance of different bass enhancement systems. Five snapshots of pop music with duration of about 10 seconds were used as the reference (unprocessed) stimuli. All the snapshots contain both transient and steady-state bass components. The processed stimuli from NLD, PV [10] and the proposed system were tested. The same high-pass filter with 120 Hz cutoff frequency was applied to all the processed stimuli, and the high-pass filtered stimuli without bass enhancement was used as the anchor. The subjects were asked to rate the stimuli from 0 (bad) to 100 (excellent).

The subjective test results from 15 subjects were normalized to the same range of [40-100], as shown in Figure 5. The NLD system achieved a high evaluation on both audio quality and bass effect. This is attributed to the fact that most subjects judged the quantity of the bass mainly on the transient, and its audio output sounds clear due to the lack of the steady-state harmonics. A reason for the low rating of the PV system is due mainly to the unnatural higher harmonics that were perceived as artifacts by some subjects. The strongest bass effect was produced by the proposed hybrid system which is attributed to the combination of transient and steady-state harmonics.

4. CONCLUSIONS

In this paper, we proposed a new psychoacoustic bass enhancement system based on the missing fundamental phenomenon, and evaluated the bass effect and audio quality of this system using subjective assessment. We have shown that the hybrid system, which combines NLD and PV, produces more impactful bass perception with acceptable distortion. In addition, the improved PV used in the proposed system results in better audio quality than the conventional PV using sum-of-sinusoids method.

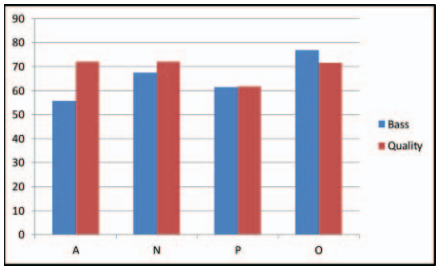

Figure 5. Subjective rating for bass enhancement systems (A = anchor, N = NLD, P = PV, O = our system).

5. ACKNOWLEDGEMENTS

The authors would like to thank Adam J. Hill from University of Essex for the use of his Virtual Bass Toolbox.

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," *Consumer Electronics, IEEE Transactions on*, vol. 47, no. 4, pp. 787–796, 2001.

[2] A. J. Hill and M. O. J. Hawksford, "A hybrid virtual bass system for optimized steady-state and transient performance," in *Computer Science and Electronic Engineering Conference (CEEC), 2010 2nd*, pp. 1–6, 2010.

[3] 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.

[4] M. R. Bai and W. Lin, "Synthesis and implementation of virtual bass system with a phase-vocoder approach," *Journal-Audio Engineering Society*, vol. 54, no. 11, p. 1077, 2006.

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

[6] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama, "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram," in *Proc. EUSIPCO*, 2008.

[7] M. Klingbeil, "Spectral Analysis, Editing, and Resynthesis: Methods and Applications", 2009.

[8] J. Laroche and M. Dolson, "New phase-vocoder techniques for pitch-shifting, harmonizing and other exotic effects," in *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pp. 91–94, 1999.

[9] K. R. Fitz and S. A. Fulop, "A unified theory of time-frequency reassignment," *Arxiv preprint arXiv:0903.3080*, 2009.

[10] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 323-332, May 1999.

[11] A. J. Hill, "Virtual Bass Toolbox." [Online]. Available: http://www.adamjhill.com/vb.html.