A TIMBRE MATCHING APPROACH TO ENHANCE AUDIO QUALITY OF PSYCHOACOUSTIC BASS ENHANCEMENT SYSTEM

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

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

ABSTRACT

Small and flat loudspeakers usually result in poor low-frequency (or bass) responses. Conventional gain equalization does not help significantly and may even result in overdriving and distortion. A psychoacoustic approach has been found to be suitable in tricking the human ear to perceive the fundamental frequency from its higher harmonics. Past research efforts have generally focused on weighting the harmonics based on the loudness matching method, but no work on timbre matching has been carried out so far. In this paper, we propose a new timbre matching technique, which can improve the sound quality of the psychoacoustically enhanced bass. This approach adjusts the amplitude of harmonics to produce similar timbre as the original audio content. Objective and subjective tests are carried out to compare the audio quality of the psychoacoustic bass enhanced signal using the equal-loudness weighting and the timbre matching methods.

Index Terms— music signal processing, timbre matching, phase vocoder, psychoacoustic bass enhancement

1. INTRODUCTION

As portable media devices are getting smaller, slimmer, and lighter, loudspeaker units that are embedded in these devices must be reduced in size and thickness. Due to the physical size limitation and cost constraint, these small loudspeaker units are unable to produce good or sufficient bass (low frequency) effect. The virtual bass system (VBS) [1]–[6] has been introduced to enhance the poor bass performance of the small and flat-panel loudspeakers by tricking the human auditory system to perceive the low-frequency component, that is not physically reproduced. The VBS is based on a psychoacoustic phenomenon known as the "missing fundamental" [7], [8], which states that higher harmonics of the fundamental frequency can produce the sensation of the fundamental frequency in the human brain. The output signal of the VBS consists of the generated higher harmonics and the original signal, so listeners can perceive the enhanced bass performance with loudspeakers having poor bass response.

VBS can be implemented in either time or frequency domain. For the time-domain approach, a nonlinear device (NLD) [2]–[5], [9]–[16] is commonly used to generate harmonics of the original signal. In the frequency-domain, a phase vocoder (PV) based VBS is introduced in [17]. However, several limitations are found with both NLD and PV methods. From previous studies [6], [18], [19], the NLD and PV methods perform complementary roles in handling different types and segments of music. NLD based VBS results in an impressive performance on the transient components such as the drum beats, while PV is more applicable to steady-state signals. Therefore, Hill *et al.* [18], [19] designed a hybrid virtual bass system. This system uses a transient content detector to handle the mixing of the NLD's and PV's outputs. Mu *et al.* [6] proposed another hybrid system with improved separation of transient and steady-state signals, and exhibits a more robust performance on different genres of music. The general framework of Mu's hybrid virtual bass system is shown in Figure 1. This system separates the input signal into transient and steady-state components using a median filter based method. The separated transient component is sent to the NLD, and the higher harmonics of the steady-state component are generated using the PV algorithm.

As shown in Figure 1, PV provides weighting control over each generated harmonic to adjust the sound quality of output. In previous research work [17]–[19], the generated harmonic components are weighted according to the equal-loudness contours. The weight values are calculated based on loudness (in phon) comparison between the fundamental components and the corresponding higher harmonics. However, to the best of authors' knowledge, there is no publication on applying timbre matching in psychoacoustic bass enhancement systems. In addition, it is found from our previous informal listening tests that the generated harmonics with the loudness matching method occasionally sounds unnatural and sharp, leading to lower audio quality.

Therefore, we propose a timbre matching method, which produces harmonics having similar timbre as the original sound source. This method is based on the fact that steady-state signal has a stable harmonic spectral structure [22], which contains the timbre information. The main motivation of this method is to improve the sound quality of

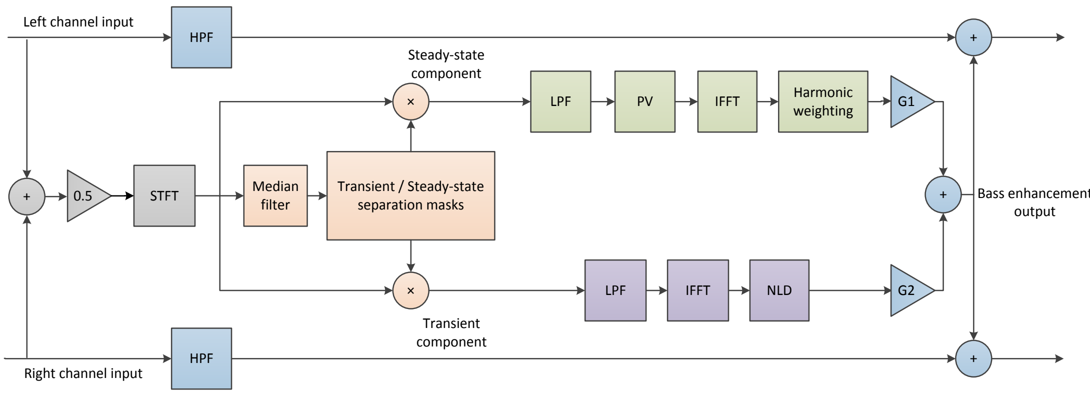

Figure 1. Framework of Mu's hybrid virtual bass system [6]. HPF: high-pass filter, STFT: short-time Fourier transform, LPF: low-pass filter, IFFT: inverse Fourier transform.

our hybrid VBS [6] by maintaining the timbre of the bass enhanced audio.

This paper is organized as follows. Sections 2 and 3 present the equal-loudness weighting and timbre matching approaches, respectively. Discussions on the objective analysis and subjective test are presented in Sections 4 and 5. Finally, our conclusion is presented in Section 6.

2. EQUAL-LOUDNESS WEIGHTING

The equal-loudness weighting method [17] is based on the equal-loudness contours, which is parameterized by

$$L_N = 4.2 + \frac{a_f(L_f - T_f)}{1 + b_f(L_f - T_f)}, \quad (1)$$

where L_N and L_f represent loudness (phon) and sound pressure level (SPL), respectively. The parameters a_f, b_f, and T_f can be fitted into polynomials, as given in [17]. Therefore, the SPL of the harmonics having the same loudness as the fundamental frequency can be estimated, and the weights of the harmonics are determined by

$$w_{dB}(i) = L_f(i) - L_f(1), \quad (2)$$

where $L_f(i)$ and $L_f(1)$ denote the SPL of ith harmonic and the fundamental frequency, respectively.

3. TIMBRE MATCHING

The proposed timbre matching is based on the source filter model, which has been widely used in speech and music synthesis [23]. The musical instrument sound can be viewed as a signal generated from a vibrating object, and then filtered by the resonance structure of the instrument [24]. In the frequency domain, the source-filter model can be illustrated as the multiplication of the vibration spectrum, which is usually modeled as a series of harmonics, with the resonance spectrum (as shown in Figure 2). The resonance spectrum contains information on timbre, and can be used to describe the timbre of a harmonic sound.

However, due to the fact that the input of the VBS is usually a piece of polyphonic music, a reliable source separation method is necessary to estimate the resonance spectrum of the bass sound source from the music. Duan *et al.* [25] introduced a source separation method for music signal based on the idea that different harmonic instruments have different harmonic structures and each structure is stable within a narrow pitch range. This is attributed to the unique and relatively flat resonance spectrum of each instrument. In the proposed VBS, the targeted low-frequency sound source is set between a narrow frequency range (around 30 to 100 Hz), so that the resonance spectrum of the desired bass harmonic sound source can be detected using the approximate invariant feature of its harmonic structure.

Based on the source-filter model introduced in [26], the resonance spectrum of the bass sound source is represented as a linear combination of fixed elementary responses:

$$h(k) = \sum_{j=1}^{J} c_j a_j(k), \quad (3)$$

where $h(k)$ represents the resonance spectrum. The weights c_j for j=1, 2, …, J determine the spectral shape of the instrument, and these weights are nearly invariant for each harmonic sound source. The elementary response $a_j(k)$ consists of a triangular filterbank uniformly distributed in the Mel-frequency scale. In the proposed VBS, the fundamental frequency is set below 100 Hz, and its corresponding 2nd to 6th harmonics are generated to produce the virtual bass effect for steady-state signals. Accordingly, we designed the filter bank covering the spectral range from around 31 to 736 Hz.

The next step is to estimate the weight c_j of the bass sound source. The spectral magnitudes in each band are summed and normalized to [0, 1]:

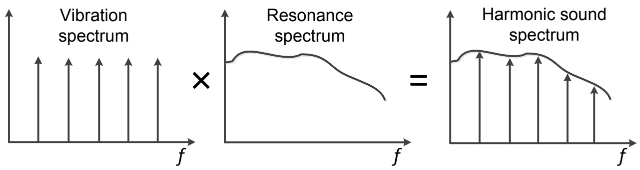

Figure 2. Source-filter model of harmonic sound generation.

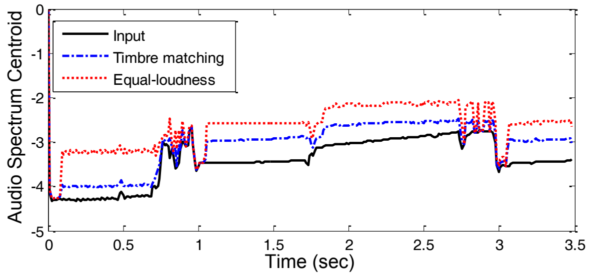

Figure 4. Audio spectrum centroid (ASC) curves of a bass guitar solo input, and the VBS-enhanced output using timbre matching method and equal-loudness weighting method.

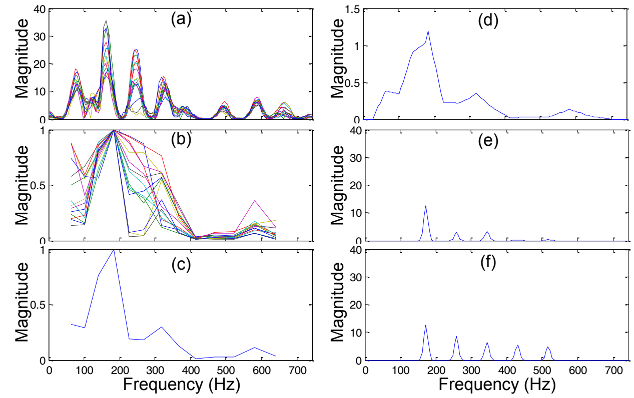

Figure 3. Plots illustrating the timbre matching processing. (a) Magnitude spectrums of the separated steady-state components in 15 successive frames. (b) Curves of b_j from (4) in 15 frames. (c) The weights c_j estimated by averaging b_j in different frames. (d) Reconstructed resonance spectrum. (e) Spectrum of synthesis harmonics using timbre matching method. (f) Spectrum of synthesis harmonics using equal-loudness weighting method.

$$b_j = \frac{\sum_{k\in j} a_j(k)\left|X(k)\right|}{\max\left(\sum_{k\in j} a_j(k)\left|X(k)\right|\right)}, \tag{4}$$

where the spectrum $X(k)$ is the separated steady-state component from the hybrid VBS [6]. Since the resonance spectrum of the bass source is different from the other sources and approximately invariant, taking an average of b_j in a number of frames can eliminate the interference and produce the estimation of the weight c_j of the bass instrument. In the proposed system, the average of b_j is calculated from 15 successive frames:

$$\hat{c}_j = \frac{1}{L}\sum_{m=0}^{L-1} b_j(n-m), \quad L = 15, \tag{5}$$

where n represents the time frame index.

Finally, the resonance spectrum $\hat{h}(k)$ is reconstructed by multiplying the estimated weight from (5) with the elementary response. To maintain the timbre at the output, the synthesis harmonics are weighted by:

$$w_{dB}(i) = 20\cdot\log\left(\frac{\hat{h}(i\cdot k_1)}{\hat{h}(k_1)}\right), \tag{6}$$

where i = 1, 2,…, 6 refer to the orders of higher harmonics, and k_1 represents the index of the fundamental frequency. Following the method in [27], the spectrum of the harmonics is synthesized by computing the main-lobe of the window transform with the appropriate frequency and magnitude. Figure 3 illustrates the process of the proposed timbre matching method. The harmonics in Figure 3(e) are weighted according to the estimated resonance spectrum; while Figure 3(f) shows the synthesis harmonics weighted using the equal-loudness method.

4. OBJECTIVE EXPERIMENTS ANALYSIS

In the objective analysis, we assess the timbre sharpness of the VBS-enhanced sound using the weighting methods with respect to loudness and timbre. Due to the additional harmonics, the average frequency of the spectrum is increased. Hence the timbre of the VBS-enhanced signal is usually perceived to be sharper than the audio input [28]. To assess the timbre sharpness level of the audio signal, we apply one of the low-level descriptors in the MPEG-7 standard known as the audio spectrum centroid (ASC) [29].

In our test case, a bass guitar solo track having a stable harmonic structure is enhanced using the PV algorithm with equal-loudness weighting and timbre matching methods. To provide a fair comparison, the weights from both methods are normalized before adjusting the harmonics. Figure 4 shows the result of the objective experiment. Both weighting methods increase the timbre sharpness of the sound, but the ASC of timbre matching method is much lower than the equal-loudness method. This result indicates that the proposed timbre matching method can generate smoother timbre than the equal-loudness weighting method.

5. SUBJECTIVE TESTING

Based on the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) approach documented in the ITU-RBS.1534-1 standard [30], we performed a subjective evaluation on the bass intensity and perceived audio quality of the proposed hybrid VBS. MUSHRA is a subjective evaluation methodology, which produces a reliable and repeatable measure of the audio signal quality [31]. In the subjective evaluation, listeners are required to evaluate

Table 1. List of processing methods for tested stimuli and their respective low-pass filters (LPF).

Index	Stimuli	LPF
1	Hidden Reference	120Hz
2	Anchor	500Hz
3	Equal-loudness with 6dB gain	120Hz
4	Timbre matching with 6dB gain	120Hz
5	Equal-loudness with 12dB gain	120Hz
6	Timbre matching with 12dB gain	120Hz

multiple VBS-enhanced audio stimuli with reference to the reference stimulus. In addition, a hidden reference (HRF) and an anchor (ARH) are included as benchmarks to assess the validity of the subjective evaluation.

Equal-loudness weighting and timbre matching methods are applied separately to the hybrid VBS [6]. The magnitudes of the harmonics are determined by:

$$M_i = M_1 \cdot 10^{(w_{dB}(i)+G1)/20}, \tag{7}$$

where M_i and M_1 represent the magnitude of ith harmonic and the fundamental frequency, respectively, and $w_{dB}(i)$ are the estimated weights derived from (2) and (6). The parameter $G1$ is the gain for PV (see Figure 1), and two values of $G1$ (6 dB and 12 dB) are adopted in our subjective evaluation.

The testing stimuli are listed in Table 1. To simulate the poor bass performance of low-end loudspeaker units, a 120 Hz high-pass-filtered stimulus was used as the reference. All the VBS-enhanced stimuli were high-pass-filtered with the same filter as the reference. A 500 Hz high-pass-filtered stimulus without bass enhancement was set as the anchor. Fifteen listeners participated in the subjective tests, and graded the stimuli in terms of bass intensity and audio quality over a score of 0-100. Bass intensity refers to the quantity of lower frequency components. When the bass of the stimulus is close to the reference, a score around 50 should be given. Scores below 50 represent “lesser bass”, and scores above 50 represent “more bass”. Audio quality is scored in terms of noise and distortion. The highest rating 100 represents a clean and completely undistorted stimulus, while the lowest rating 0 represents an extremely distorted and unacceptable stimulus

The test scores are plotted in Figure 5. As expected, the hidden reference results in a bass intensity around 50 points and the best quality scores, while the anchor received the worst evaluation in the bass intensity. Both equal-loudness weighting and timbre matching methods with the gain of 6dB exhibit very limited bass enhancement effect as compared with the reference. The limited bass enhancement is attibuted to the small gain of 6dB, which is insufficient to produce obvious virtual bass enhancement. Increasing the gain can improve the virtual bass enhancement but might

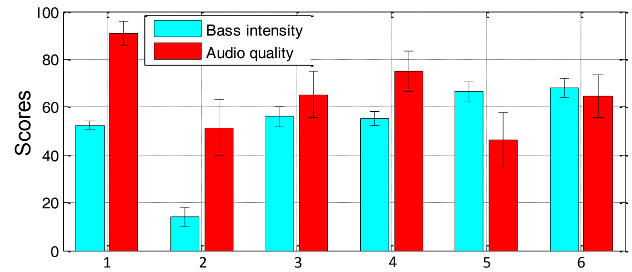

Figure 5. Average subjective scores in bass intensity (blue bar) and audio quality (red bar) with 95% confidence intervals for VBS.

lead to higher distortion in the virtual-bass enhanced audio. As shown in Figure 5, the equal-loudness method having a gain of 12dB shows stronger bass effect as well as relatively low rating for its audio quality. Although the proposed method also suffers from the audio distortion with increasing gain, its audio quality is still acceptable and significantly better than the equal-loudness weighting method. At the same bass intensity, the timbre matching method produces a better audio quality than the equal-loudness weighting method.

6. CONCLUSION

A new psychoacoustic bass enhancement with timbre matching, as well as its effectiveness in generating a more natural bass effect were discussed in this paper. Based on the psychoacoustic theory, this bass enhancement system produces a virtual bass enhancement effect without excessive bass boosting of the loudspeaker unit. A new timbre matching method was designed to preserve the timbre of bass sound at the output, while prior work only focused on matching the loudness attribute. The objective results indicated that the proposed method can improve the timbre sharpness problem of the psychoacoustic bass enhancement system. In the subjective assessment, the timbre matching method was found to produce an improved audio quality (10-18 points) compared to the conventional equal-loudness weighting method.

7. ACKNOWLEDGEMENTS

The authors would like to thank Adam J. Hill from University of Essex for the use of his Virtual Bass Toolbox. This work is also partially supported by the Singapore Ministry of Education Academic Research Fund Tier-2, under research grant MOE2010-T2-2-040.

8. 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," 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] H. Mu, W. S. Gan, and E. L. Tan, "A psychoacoustic bass enhancement system with improved transient and steady-state performance," in *37th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 2012, pp. 141 –144.

[7] J. F. Schouten, "The residue and the mechanism of hearing," in *Proceedings of the Koninklijke. Nederlandse Akademie van Wetenschappen*, 1940, vol. 43, pp. 991–999.

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

[9] M. Shashoua and D. Glotter, "Method and system for enhancing quality of sound signal," U.S. Patent 593037327-Jul-1999.

[10] D. Ben-Tzur and M. Colloms, "The effect of MaxxBass psychoacoustic bass enhancement on loudspeaker design," in *106th Convention of the Audio Engineering Society*, Munich, Germany, 1999.

[11] G. F. M. D. Poortere, C. M. Polisset, and R. M. Aarts, "Ultra bass," U.S. Patent 613433017-Oct-2000.

[12] E. R. Larsen and R. M. Aarts, "Reproducing low-pitched signals through small loudspeakers," *Journal of the Audio Engineering Society*, vol. 50, no. 3, pp. 147–164, 2002.

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

[14] B. Pueo, G. Ramos, and J. J. Lopez, "Strategies for bass enhancement in Multiactuator Panels for Wave Field Synthesis," *Applied Acoustics*, vol. 71, no. 8, pp. 722–730, 2010.

[15] L. K. Chiu, D. V. Anderson, and B. Hoomes, "Audio output enhancement algorithms for piezoelectric loudspeakers," in *IEEE Signal Processing Society 14th DSP Workshop & 6th SPE Workshop*, Sedona, Arizona, 2011, pp. 317–320.

[16] T. Holman, "Low-frequency range extension and protection system for loudspeakers," U.S. Patent 801908813-Sep-2011.

[17] M. R. Bai and W. Lin, "Synthesis and Implementation of Virtual Bass System with a Phase-Vocoder Approach," *Journal of the Audio Engineering Society*, vol. 54, no. 11, pp. 1077–1091, 2006.

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

[19] A. J. Hill and M. O. J. Hawksford, "Wide-area psychoacoustic correction for problematic room modes using non-linear bass synthesis," *Journal of the Audio Engineering Society*, 2012.

[20] F. A. Karnapi and G. Woon-Seng, "Method to Enhance Low Frequency Perception from a Parametric Array Loudspeaker," *Watermark*, vol. 1, 2002.

[21] J. Choi, J. Kim, Y. Kim, and S. Ko, "Sound enhancement apparatus and method," U.S. Patent EP233410315-Jun-2011.

[22] 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.

[23] V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, "Discrete-time modelling of musical instruments," *Reports on progress in physics*, vol. 69, no. 1, p. 1, 2006.

[24] M. Muller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal Processing for Music Analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, pp. 1088–1110, Oct. 2011.

[25] Z. Duan, Y. Zhang, C. Zhang, and Z. Shi, "Unsupervised Single-Channel Music Source Separation by Average Harmonic Structure Modeling," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 766–778, May 2008.

[26] T. Heittola, A. Klapuri, and T. Virtanen, "Musical instrument recognition in polyphonic audio using source-filter model for sound separation," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2009, pp. 327–332.

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

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

[29] H.-G. Kim, N. Moreau, and T. Sikora, *MPEG-7 audio and beyond: audio content indexing and retrieval*. Wiley, 2005.

[30] ITU-R BS.1534-1, "Methods for the subjective assessment of small impairments in audio systems," *International Telecommunications Union, Geneva, Switzerland*, 2001.

[31] G. Stoll and F. Kozamernik, "EBU Listening Test on Internet Audio Codecs," *EBU Technical Review*, no. 283, Jun. 2000.