Virtual Bass System by Exploiting the Rhythmic Contents in Music

Siby Charley Pulikottil

Samsung R&D Institute - Bangalore
Bangalore, India

siby.c@samsung.com

***Abstract*— Current breed of low to mid end speakers and earphones are unable to recreate the bass frequency range satisfactorily. Bass enhancement using the psychoacoustic principle of missing fundamental is particularly suitable for such scenarios. Traditionally there are two techniques employed for this such as Non Linear Devices (NLD) and Phase Vocoders (PV). However each one comes with its own set of cons namely higher distortion in NLD and transient smearing in PV respectively. With the proposed approach we are trying to create a Virtual bass system that can leverage the rhythmic information that is normally associated with the lower frequency region of a music signal. By applying the proposed approach the intermodulation distortion can be greatly reduced (~ 56%) without affecting the bass quality.**

Keywords—virtual bass, audio, signal processing, nonlinear processing, rhythm, distortion, peak picking

I. Introduction

With the advances in the field of audio processing, the playback scenario changed dramatically over the last decade. We are exposed to different playback configurations and high end speaker systems. But the common low to mid segmented speakers/earphones fail to recreate the bass level frequencies typically in the ranges below 120Hz. In such conditions perception of bass frequencies can be created through psychoacoustic bass enhancement by applying the phenomena called “Missing Fundamental”. There are basically two common approaches towards this concept and they are Non Linear Devices (NLD) and Phase Vocoder (PV) respectively. Now in the proposed approach using the beat detection we are trying to leverage on the rhythm associated with the lower frequencies thus overcoming the drawbacks associated with both NLD and PV.

A. Missing Fundamnetal

This phenomenon is the result of the complex pitch extraction procedure happening in the inner ear and brain. When subjected with a complex signal, brain and ear tries to perceive the pitch by linking various spectral components to one another [1]. When spectral components are equiv spaced the mechanism takes in the greatest common factor among frequencies as the pitch. 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. We are applying the same principle in virtual bass system, where the lower frequency pitches are simulated through their upper harmonics.

B. Non Linear Devices(NLD)

This is the most commonly found technique of harmonic generation used in virtual bass systems and is widely used because of its ease to implement and faster execution. The core idea in this method is to make the signal pass through a non linear device. The applied non linearity causes the harmonics along with some amount of inter modulation distortion (IMD) and this IMD happens to be the main disadvantage of NLD. A typical NLD uses a polynomial approximation function as defined in (1).

$$y = h_1 x + h_2 x^2 + ... + h_n x^n \tag{1}$$

where *h* is a vector containing N polynomial coefficients with *x* and *y* representing input and output respectively. The NLD is a memory less system and current output depends only on the input. Here the main source of IMD is the time domain operation where the entire signal is applied with non-linearity creating unwanted frequency combinations.

C. Phase Vocoder(PV)

Recently an alternative to the NLD was introduced using Phase Vocoder (PV) as harmonic generator [2]. The main advantage of this method is the removal of inter modulation distortion (IMD) by doing frequency domain operations. Apart from that this approach also provides better harmonic control over the output. In this method the input is initially divided into overlapping frames typically of 50-250msec. Now the Fast Fourier Transform (FFT) is applied on these time domain frames and the required pitch shifting based operations are done by keeping the phase coherence. Finally the resultant time domain signal is retrieved by the inverse of FFT.

Although PV based approach gets rid of the IMD, it has its own disadvantage. Frequency domain approach towards PV often results in having high frequency resolution at the cost of lower time resolution. This results in the smearing of transients out of the system and causes loss of percussion in subjective perception. Hence the PV approach is normally more suitable for steady state signals.

978-1-4799-1823-2/15/$31.00 ©2015 IEEE

Combining together the advantages of NLD and PV for better system resulted in a hybrid approach [3]. And this technique uses a transient detector to switch between the NLD and PV.

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

II. PROPOSED APPROACH

It's well-known that although NLD causes the IMD, the amount of upper harmonics and in turn the bass effect is more profound in NLD approach compared to PV. Although it's the non-linearity which creates IMD, the real problem here is the usage of the complete signal against selective frequency usage as in PV.

The basic thought behind this proposed approach came from the need to create a virtual bass system that will reduce IMD as well as preserve transient percussions. In the case of any musical signal the low frequency portions predominantly gives perception of rhythm and mid/high frequency regions gives the melodic information.

In such a scenario the end user only perceives the rhythm of the music through the bass frequencies. The rhythm in turn is perceived as the beats of a signal [4]. Hence, the entire problem can be reduced to the identification of correct beat positions followed with the calculation of prominent frequencies (only a few) during the beats. Thus instead of applying the entire signal over NLD, we can apply those prominent frequencies alone resulting in an effective reduction in IMD. The detailed system block diagram is shown in (Fig.1). Currently the system is implemented for mono channel, but it can be easily scaled to dual channel.

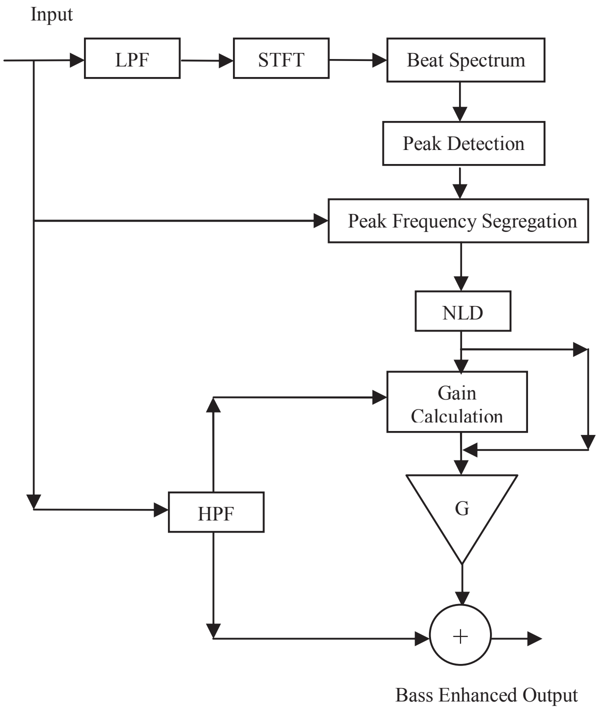

Fig. 1. Proposed Virtual Bass System

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

A. Filtering(LPF & HPF)

The bass enhancement algorithm needs to operate only on the low/bass frequency (120Hz) region. Hence the input signal is allowed to pass through a low pass filter (LPF). Since the cut off has to be sharp, a 4th order elliptic FIR filter was chosen with a cutoff frequency of 120Hz. The higher frequency portions of the input signal have to be retained for final summing with the bass enhanced signal. So to retain the high frequency region the input signal is made to pass through a high pass filter (HPF).

B. Short Time Fourier Ttransform(STFT)

For finding the beat spectrum initially the input is converted into the frequency domain. STFT [5] gives a Time-Frequency representation which assumes the signal to be stationary for a narrow time frame. In this case the signal is initially framed and windowed at 46.4msec at a sampling frequency of 44.1 KHz. Overlapping hamming windows are used with a 50% overlap. The N =2048 point FFT is applied on each time frame window, resulting in the completed Time-Frequency representation. Now the complete STFT matrix is divided into blocks along the time axis with each 30 frames constituting as a block. This is done to give fine-tuned resolution for calculation of beat spectrum in the consequent step of the algorithm. This blocking gives a fine resolution of 0.696sec of unique data for aforementioned sampling frequency.

C. Beat Spectrum

To exploit the rhythmic nature of the music the beat information of the signal has to be analyzed. And to find the beat spectrum [6] auto correlation method is used. Here all the operations are carried out in a block loop for the optimum resolution. For finding the beat spectrum the squared magnitude of FFT spectra corresponding to each time frame is found from the STFT matrix. Each of the frequency columns across the frames are analyzed for finding the repetitive pattern. To do this, the auto correlation is applied on each frequency column in the transposed STFT matrix resulting in output (Au). The auto correlation is carried out using FFT/IFFT procedure by discarding symmetric half. The Au matrix is analyzed using (2) and (3) to identify the frame corresponding to maximum similarity and is defined as the strongest repeating period.

$$BS(i) = \frac{1}{n}\sum_{j=1}^{n} Au(i, j) \tag{2}$$

$$RP = \max(BS) \tag{3}$$

where *i = 1 to number of frames in a block, j = frequency bins*, and *n = N/2+1*. The relation in (2) shows the beat spectrum (BS) calculation for a block of 30 frames. For finding the

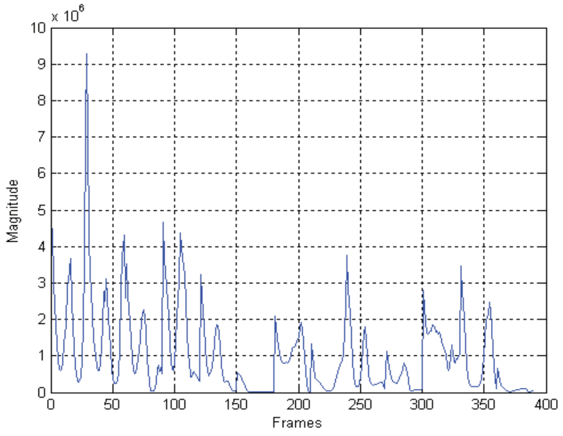

Fig. 2. Beat Spectrum

strongest repeating period (RP), the first coefficient (time lag of 0) of BS is discarded. Fig.2. shows a typical beat spectrum plot for a complete signal. Once the time frame with highest repetition is found, the next objective is to find the dominant frequency component out of that particular time frame. In order to do that, the input signal for that frame is taken and dominant frequency out of it is found using the autocorrelation method. Here the dominant frequency along with its strength is required for the efficient peak detection. So the auto correlated peak along with magnitude is saved for each block, resulting in block number of peak frequencies with their peak magnitudes.

D. Peak Detection

To find the most prominent N peaks out of all the peak frequencies saved during block iterations, we have devised a unique peak detection procedure. The following are the hierarchical steps used for the peak detection process.

- Initially all the frequency peaks are subjected for removing the repeating values.
- Now all the remaining values are sorted in the descending order of its magnitude.
- The sorted result is analyzed for peaks within a tolerance band of +/- 5Hz on either side of the peak.
- Based on the above step near peaks are approximated with the strongest one in their vicinity.
- Again the step 2 is repeated to remove redundant values with a single one.

Finally out of the remaining data (already magnitude sorted) the top N frequency points are selected. During our implementation of the system we choose N = 4, thus resulting in four frequency points.

E. Peak Frequency Segregation

As per the proposed algorithm, the IMD is reduced by using selective prominent frequency content instead of the entire signal. In order to facilitate that, the most impactful frequency points obtained from peak detector has to be separated out from the complete signal. So the input signal is notch filtered for these N frequencies and each of these notched results is subtracted from the input signal. Thus N signal vectors are obtained with each one predominantly having only that particular frequency content. The notch filter followed by subtraction method is favored instead of peak filtering due to the less distortion of notch compared to peak. Here we are using a 2nd order IIR filter for notching.

F. Non Linear Device (NLD)

Having found the peak frequencies and separated them out of the input signal, the algorithm goes back to the time domain operations. From the study of different nonlinear devices [7], Arc Tangent Square Root (ATSR) performs well resulting in good bass both in baseline and transients. Here the segregated input signals are allowed to pass through ASTR sequentially. The resultants are then made to combine back as a single bass enhanced output signal.

G. Gain Calculation

After creating the bass enhanced output it needs to be recombined with the higher frequency region which was separated at the beginning. Due to processing of the low frequency region there is obvious gain difference between the

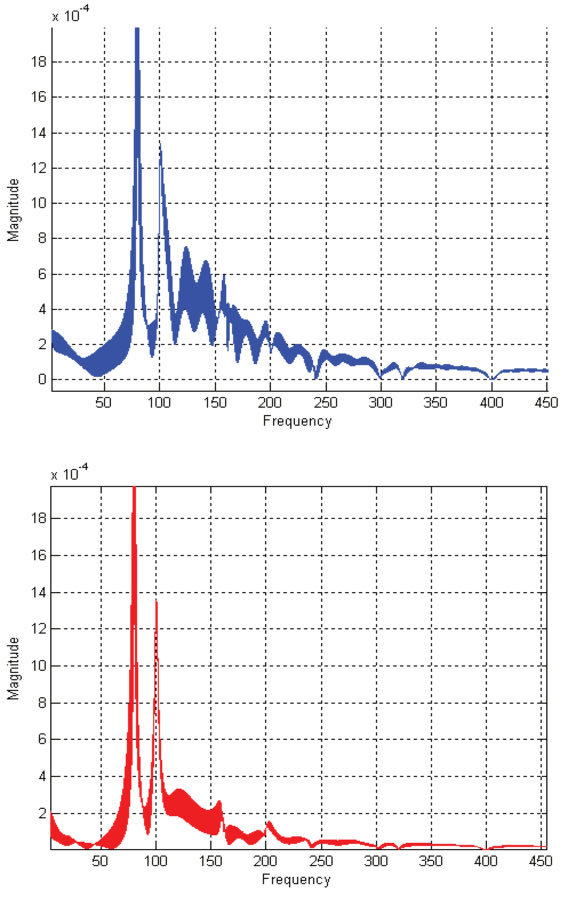

Fig. 3. Distortion Spectra from (a) NLD (b) Proposed Approach

low and high frequency regions. So in order to iron out this level difference each region is separately analyzed and suitable gain value is calculated and applied on to the low frequency region. The gain scaled low frequency region is then combined along with high frequency resulting in the final bass enhanced output signal.

III. SIMULATION & TESTING

For validation of the algorithm we have adopted a dual testing strategy such as 2Tone Distortion test and MUSHRA [6] based subjective testing. By this, initially we used a dual tone signal as input to the system to show the reduction of distortion. Later the system is subjected to MUSHRA based subjective listening test.

A. 2Tone Distortion Test

According to the proposed algorithm, the usage of selective frequency components will reduce the amount of distortion created in the final output. Now to test the validity of this, we created a 2Tone signal with 80 Hz and 100 Hz. The signal was allowed to pass through the proposed approach and a normal NLD based approach. The input 2Tone signal consisted of 10 sec of data. For this distortion analysis, final output is filtered off the input frequencies along with its entire harmonics up to 6 levels in both the methods to get the plotted waveform. After removing these harmonics, the high frequency region of the output should be ideally similar to the high pass filtered input signal. Any extra components in this section are caused predominantly due to the IMD. Fig.3. shows the waveform plot with visible change in the removal of additional peaks with the proposed method. Now the mean energy of the high pass region starting from 120 Hz was calculated for both methods. These values were again normalized with the mean energy of the high pass filtered input resulting in 2.7608(NLD) and 1.2109(Proposed approach). Thus it can be seen that with the proposed approach the unwanted energy caused by IMD can be reduced approximately by 56% with respect to the NLD.

B. MUSHRA based Perceptual testing

MUSHRA stands for Multiple Stimuli Hidden Reference and Anchor. To apply this method for perceptual testing we have identified 5 test vectors with all of them being music signals and most of them bearing an inherent baseline. Since we are comparing our proposed approach against both NLD and PV, we have chosen most of the test vectors with good amount of transient contents. For the testing procedure all the 5 test vectors with each having 10 sec were processed using our proposed approach, NLD and PV. We have also created the anchor for each of the input vector as a high pass filtered output. Now all the outputs along with the anchor are subjected for perceptual testing among 6 people in a hidden fashion. Fig.4. shows the bar graph explaining the results of the subjective testing with T1, T2 etc denoting the test vectors. During the perceptual test each of the subjects are asked to score the streams in a scale of 10 – 90 with 10 being minimum and 90 being maximum. All scores have to be in multiples of 10. All the subjects are asked to give the scores based on the amount and quality of bass. For each test stream, the average score is taken against different approaches. From Fig.4 it is evident that for almost all input cases the proposed approach outperforms the other methods. Since most of the test cases were having transient contents, the PV performed poorer than NLD in most cases.

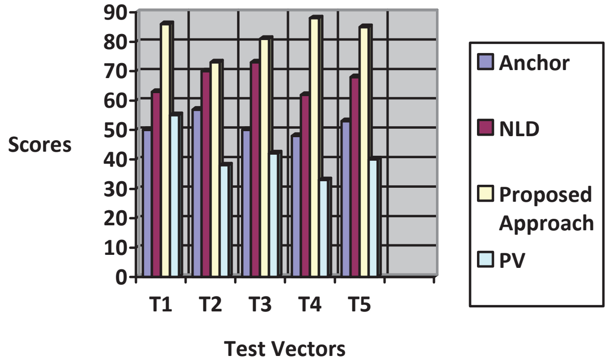

Fig. 4. Chart showing Subjective Perceptual Scores

IV. CONCLUSION

In this paper we proposed a novel virtual bass technique utilizing the rhythmic information in the low frequency part of the signal along with the normal NLD. Here the rhythm information is extracted through the calculation of beat spectrum. The results obtained have validated the IMD reduction (~56%) along with greater bass quality through subjective listening test. Thus we have shown that the proposed method successfully overcomes the major drawbacks of IMD and transient smearing associated with NLD and PV respectively.

ACKNOWLEDGMENT

I as an author would like to thank Mr. Adam J Hill for the usage of his Virtual Bass Toolbox for the generation of NLD and PV outputs for testing purpose.

REFERENCES

[1] Schouten, J.F, R.J. Ritsma, B. Lopes Cardozo. “Pitch of the residue.” Journal of the Acoustical Society of America, Volume 34, Number 8, pp. 1418-1424, September 1962.

[2] Bai, M.R, W.C. Lin. “Sythesis and implementation of virtual bass system with a phase-vocoder approach.” Journal of the Audio Engineering Society, Volume 54, Number 11, pp. 1077-1091. November 2006.

[3] A.J.Hill, M.O.J.Hawksford “A Hubrid Virtual Bass Sytem for optimized steady state and transient performance” in Computer Science and Electronic Engineering Conference(CEEC), 2010 2nd, 2010, pp 1-6

[4] Foote.J, FX Palo Alto Laboratory Inc.; Uchihashi, S. “The beat spectrum: a new approach to rhythm analysis” in IEEE International Conference on Multimedia and Expo, 2001. ICME 2001, Tokyo, Japan, 22-25 Aug, 2001.

[5] Allen, J.B. ; Bell Laboratories, Murray Hill, NJ ; Rabiner, L "A unified approach to short-time Fourier analysis and synthesis" Proceedings of the IEEE (Volume:65, Issue:11), Nov. 1977, Pages: 1558-1564.

[6] Z.Rafi, B.Prado. "A Simple Music/Voice separation method based on the extraction of repeating musical structure", 36th International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, May 22-27, 2011.

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

[8] Mason, A.J. "The MUSHRA audio subjective test method." BBC Research & Development White Paper, WHP 038, September 2002.