A hybrid virtual bass system for optimized steady-state and transient performance

Adam J. Hill and Malcolm O. J. Hawksford
Audio Research Laboratory
School of Computer Science & Electronic Engineering, University of Essex
Colchester, United Kingdom
ajhilla@essex.ac.uk mjh@essex.ac.uk

Abstract— **Bandwidth extension of a constrained loudspeaker system is regularly achieved employing nonlinear bass synthesis. The method operates on the doctrine of the missing fundamental whereby humans infer the presence of a fundamental tone when presented with a signal consisting of higher harmonics of said tone. Nonlinear devices and phase vocoders are commonly used for signal generation; both exhibiting deficiencies. A system is proposed where the two approaches are used in tandem via a mixing algorithm to suppress these deficiencies. Mixing is performed by signal transient content analysis in the frequency domain using constant-Q transforms. The hybrid approach is rated subjectively against various nonlinear device and phase vocoder techniques using the MUSHRA test method.**

Keywords-virtual bass, audio, psychoacoustics, nonlinear signal processing, phase vocoder, transient detection

I. INTRODUCTION

Loudspeaker systems commonly operate over a restricted bandwidth due to design compromises, thus prohibiting them from reproducing the full bandwidth of the source signal. A typical solution to this problem is based on the concept of the "missing fundamental" whereby a set of higher harmonics of a fundamental frequency causes a listener to infer the presence of the fundamental frequency although it is not physically present. Nonlinear processing is applied to the input signal in the form of a virtual bass algorithm to enhance the perception of the lower-frequency range. This methodology is of great use to systems with small drive units that cannot efficiently reproduce low-frequencies.

Nearly all virtual bass algorithms utilize either a nonlinear device (NLD) or a phase vocoder (PV). The NLD approach operates entirely in the time domain and introduces harmonic distortion based on the specific nonlinear device chosen. This approach results in positive subjective results for transient signal components, such as drum beats, where the input signal is spectrally-rich and is largely insensitive to harmonic distortion. The NLD approach does not allow for accurate control over the specific harmonic components, however, causing an unnatural virtual bass effect for pitched signal components. This unnatural characteristic is due to the higher-harmonic components which introduce a metallic quality to the sound. In addition, NLD's are highly input amplitude sensitive. Low magnitude signals will result in nearly no virtual bass effect, causing a noticeable inequality in the effect over a dynamic signal's duration.

The PV approach, on the other hand, operates in the frequency domain, which allows for precise control over the individual harmonic components and is input amplitude insensitive, giving a subjectively-equal virtual bass effect for both soft and loud signals. A drawback to the PV approach, though, is that it requires a sufficiently large analysis window in the time domain to achieve enough low-frequency resolution to avoid intermodulation distortion. This leads to a smearing of transient signal components, resulting in an unnatural effect.

A virtual bass algorithm has been developed that draws on the transient handling capabilities of an NLD while also taking advantage of the steady pitched component and amplitude-insensitive operation of a PV. The remainder of this paper will present NLD and PV virtual bass methodologies followed by a description of the mixing algorithm utilized within the hybrid system. Subjective test results will be presented with a discussion on the overall effectiveness of the proposed virtual bass algorithm.

II. THE MISSING FUNDAMENAL

The phenomenon of the missing fundamental, or the residue pitch, is a result of the complex pitch-extraction mechanism within the inner ear and the brain. When presented with a spectrally-complex sound, the pitch extraction mechanism attempts to make sense of the received signal by relating various spectral components to one another [1]. When spectral components are equally spaced from one another, this will result in a perceived pitch corresponding to the greatest common factor of the frequency values (in Hz) that falls within the audible range of 20 Hz – 20 kHz. For instance, if the source contains spectral components at 200, 300, 400 and 500 Hz the overall perception will correspond to a harmonically-rich tone at 100 Hz.

This effect can operate using only two higher harmonic components of the fundamental. Adding additional harmonics will increase the sharpness of the signal timbre (sound quality) as the average frequency of the components increases [1].

When applying the missing fundamental for low-frequency applications, it is important to keep the average frequency of

978-1-4244-9030-1/10/$26.00 ©2010 IEEE

all spectral components to a minimum so that the perceived pitch is as close in timbre to the fundamental as possible. Minimizing the amount of harmonic components introduced will also preserve the fidelity of the source signal since these virtual bass components are a form of distortion which should ideally be kept to a minimum.

III. Nonlinear Device Virtual Bass

A nonlinear device (NLD) is the most common harmonic generator implemented within virtual bass systems for a number of reasons. First, the NLD is memoryless, allowing for real-time applications. NLDs generally operate using a polynomial approximation of a chosen function. The calculated coefficients are then applied to the input signal as defined in (1).

$$y = \sum_{i=0}^{N} h_i x^i \tag{1}$$

where, h is a vector containing the N polynomial coefficients with x and y representing the signal input and output, respectively [2].

The NLD virtual bass technique operates in the time domain, applying the effect over all spectral components of the signal. This often introduces intermodulation distortion to the signal if there are two closely-spaced spectral components in the input signal. While it has been argued that these components cause minimal auditory artifacts due to psychoacoustical masking at the Basilar membrane in the inner ear [2], intermodulation distortion is an unwanted peripheral to the NLD virtual bass system, which must be handled carefully.

Early virtual bass research utilized a full-wave rectifier (FWR) for the NLD [3]. The FWR can be very simple to implement, but suffers from the fact that it generates only even-order harmonics. A FWR applied to a 100 Hz pure tone would result in harmonic distortion introduced at 200, 400, 600 Hz and so on. Following the principle of the missing fundamental, this harmonic series should result in a perceived pitch of 200 Hz rather than 100 Hz. The perceived pitch is a full octave higher than the target pitch perception which results in an inaccurate virtual bass effect.

This problem has led to a significant body of research to develop the ideal NLD for virtual bass applications. A wide range of NLDs are presented in [2], where they are each objectively and subjectively evaluated to best judge performance. The second exponential-type NLD in [2] was rated highly in both objective and subjective tests and was therefore chosen as the NLD for this work. The input-output relationship is shown in Fig. 1.

NLD virtual bass systems are implemented with a series of filters to give approximate control of the spectral components of the effect. The input signal is first processed by a low-pass filter (LPF) with a cutoff frequency set to the upper limit of the required low-frequency extension. This low-passed signal is then processed by the NLD, generating the harmonic components.

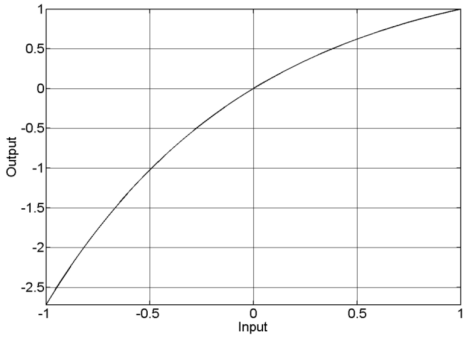

Figure 1. Input-ouput relationship for exponential NLD

Next, the NLD output is sent through a bandpass filter (BPF) to remove the fundamental spectral components and to roughly shape the harmonic components. If only a low-frequency boost is required (as opposed to a bandwidth extension), the BPF can be replaced by a LPF. After the BPF, gain is applied to the signal and then combined with a delayed version of the original signal. The overall NLD virtual bass process is shown in Fig. 2.

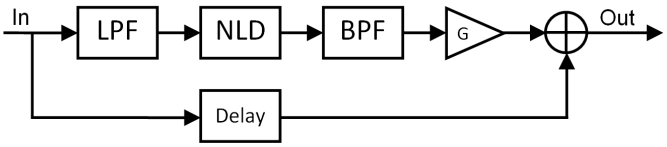

Figure 2. NLD virtual bass procedure

A widely-utilized commercial NLD-based virtual bass system is called MaxxBass [4]. In addition to the system architecture in Fig. 2, MaxxBass uses equal-loudness processing to provide a virtual bass effect subjectively equal in level to the unprocessed signal.

IV. Phase Vocoder Virtual Bass

An alternative to the NLD virtual bass approach has emerged in recent years utilizing a phase vocoder (PV) as the harmonic generator [5]. The PV virtual bass approach provides superior harmonic control, allowing for selective harmonic inclusion in the effect. Since this approach operates in the frequency domain, the intermodulation distortion can be effectively avoided, unlike with NLDs.

PVs operate by splitting an input signal into short time-domain windows (generally between 50 – 250 ms). The PV takes the fast Fourier transform (FFT) of each time window, applies the required processing while maintaining phase coherence and then generates the output signal either by sum-of-sinusoids or inverse Fourier transforms where each window is overlap-added to minimize amplitude-modulation effects. This present work utilizes the sum-of-sinusoids method.

A disadvantage to the PV arises due to the trade-off between time and frequency resolution. Virtual bass systems require adequate frequency resolution to allow for accurate harmonic

generation in addition to avoiding intermodulation distortion. Frequency resolution can be determined by (2).

$$f_{res} = {}^{1}/_{t_w} \tag{2}$$

where, f_{res} is the frequency resolution (Hz) and t_w is the window length (s). For example, a 125 ms window gives 8 Hz resolution while a 500 ms window gives 2 Hz. This issue leads to smeared transient performance which is clearly evident when applied to audio signals such as drum beats.

Previous solutions to this problem have involved reinitializing the phase within the algorithm when a transient is encountered [6] and also removing any transients from the input signal and then reinserting them, unprocessed, at the PV output [7]. The phase re-initialization solution can prove difficult as it relies on precise transient detection; otherwise, phase re-initialization will occur in excess causing poor phase coherency for the steady-state signal components. The transient removal method has had low ranking in subjective tests since transient signal components are not addressed within the effect [7].

Even through the PV cannot handle transients perfectly it does perform well on pitched signal components. Unlike the NLD system, PV virtual bass does not require a LPF on the input stage, as the algorithm can selectively apply the effect to frequency bins. Within the PV the selected frequencies are pitch shifted to the desired harmonic frequencies and amplitude adjusted to match any equal-loudness requirements; therefore no BPF or HPF is necessary on the output stage.

Since the PV virtual bass system is more computationally demanding, it is necessary to down-sample the input signal for real-time applications. This requires a LPF before the down-sampling process to avoid any spectral aliasing. Once the signal has been processed, it can be up-sampled to the original sampling rate and recombined with the delayed original signal. The overall PV virtual bass process is shown in Fig. 3.

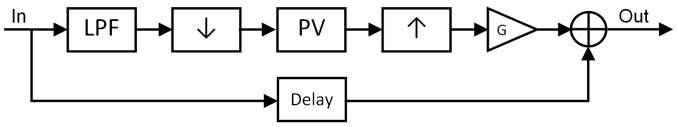

Figure 3. Phase vocoder virtual bass procedure

While PVs are commonly used for audio effects such as pitch shifting and time stretching [8], there are no known commercial applications for PV virtual bass.

V. Hybrid Virtual Bass

A virtual bass system that exploits the respective strengths of the NLD and PV systems but circumvents their weaknesses should provide a bass synthesis less sensitive to changes in input signal content. When the input signal has a high transient content, the system favors the NLD output and conversely, when the signal is more pitched the PV effect is utilized.

This hybrid approach requires a transient content detector (TCD) that analyzes successive time domain windows of the input signal and appropriately weights the respective virtual bass algorithms that are running in parallel (Fig. 4).

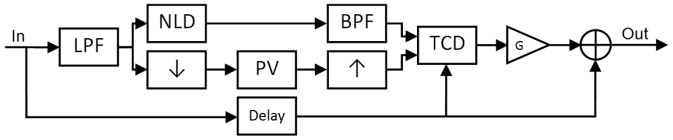

Figure 4. Hybrid virtual bass procedure

A. Constant-Q Transform

The TCD needs a time-frequency transform capable of sufficient low-frequency resolution to accurately identify the transient and pitched signal components. Discrete Fourier transforms (DFT) are commonly used for this purpose, but suffer from poor low-frequency resolution since the transform operates over linearly spaced frequency bins. This lack of spectral definition can inhibit the detection of differences between transient and pitch signals in the low-frequency range (below 100 Hz); therefore it is necessary to utilize a transform which provides higher resolution.

A constant-Q transform (CQT) is a powerful tool used for musical signal analysis. The advantage of the CQT is that it provides a constant ratio of center frequency to resolution. DFTs maintain equal resolution through all linearly spaced frequency bins, but CQTs give frequency-dependent resolution based on logarithmically spaced frequency bins [9]. This is achieved by scaling the time domain analysis window length based on the frequency bin under inspection (3), allowing for CQT calculation through (4).

$$N_k = {}^{f_s Q}/_{f_k} \tag{3}$$

$$X_k = \frac{1}{N_k} \sum_{n=0}^{N_k-1} w_k(n) x(n) e^{\frac{-j2\pi Q n}{N_k}} \tag{4}$$

where, N_k is the analysis window length for frequency bin k, f_s is the system sampling rate (Hz), Q is the ratio of frequency to resolution, f_k is the frequency bin value (Hz) and X_k is the CQT for frequency bin k, with w_k representing the analysis window function applied to input signal, x [9]. The advantage of CQT over DFT is shown in Fig. 5 where a CQT is directly compared to a DFT with the same maximum window size for a signal consisting of pure tones from 20 – 100 Hz (10 Hz increments) and 100 – 1000 Hz (100 Hz increments).

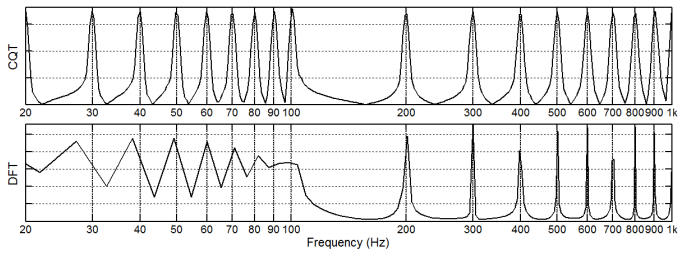

Figure 5. CQT (top) and DFT (bottom) comparison for signal with sinusoidal components at each grid line (equal maximum analysis window size for both transform methods)

The CQT and DFT comparison highlights a tradeoff between the two methods. Since the DFT has fixed analysis window size, the frequency resolution will be equal (in Hz) over all frequency bins; therefore, on a logarithmic scale (as in Fig. 5) there is higher resolution as frequency increases, but below 100 Hz there are significant inaccuracies. The CQT method, while not as accurate at higher frequencies, maintains a constant ratio of frequency resolution to give equal accuracy on a logarithmic scale. This results in much more accurate spectral transforms for frequencies below 100 Hz due to the CQT's adaptive analysis window size. The improved accuracy of the CQT will allow for detailed analysis within the TCD to best track the dynamic nature of an audio signal.

B. Transient Content Detector

The transient content detector (TCD) operates in the frequency domain, utilizing the improved low-frequency resolution of the CQT. The change in spectral energy content is tracked between successive analysis windows by targeting frequency bins in the virtual bass range (below ~100 Hz). When the change in energy exceeds a certain threshold, the overall weighting function is incremented in the NLD direction, while moving in the PV direction, otherwise.

A dynamic time window was applied to a continuous 60 Hz sinusoid to test the performance of the TCD. The window was constructed to replicate alternating transient and pitched components. Ideally, the TCD should heavily weight the NLD when a transient is encountered and then quickly shift to the PV, while the TCD should remain at full PV weighting for the pitched signal components (Fig. 6).

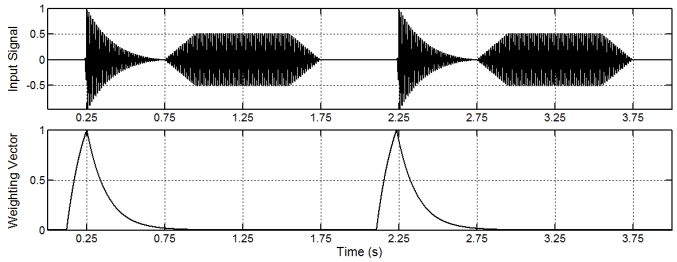

Figure 6. TCD weighting function (bottom) due to input signal (top)
(0 = full PV weighting, 1 = full NLD weighting)

The test shown in Fig. 6 highlights the TCD's functionality as required for the hybrid virtual bass system. The NLD is used for transient components while the PV is used for the steady, pitched parts. The smooth transition between operating states ensures no additional distortion will be introduced from this automatic mixing algorithm.

C. Musical signal performance

Although the TCD performs as expected with the synthesized test signals, it is necessary to perform further testing with real-world musical signals to ensure the algorithm functions appropriately within the hybrid virtual bass system.

Two clips of music were chosen to test the TCD: one consisting of a sparse drum beat (Fig. 7) and the other containing slow, low-pitched vocals (Fig. 8).

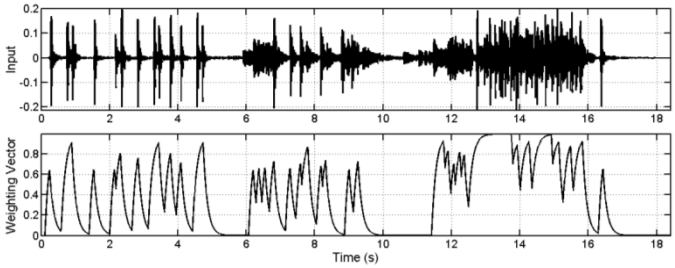

Figure 7. TCD weighting function (bottom) due to drum beat sample (top)
(0 = full PV weighting, 1 = full NLD weighting)

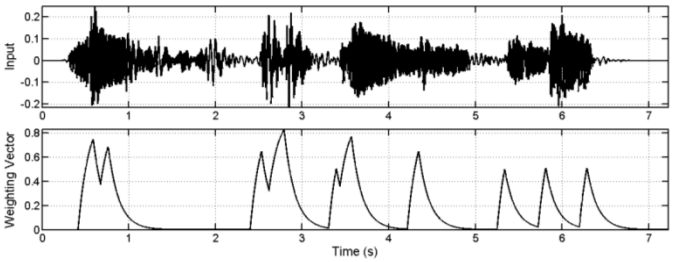

Figure 8. TCD weighting function (bottom) due to vocals sample (top)
(0 = full PV weighting, 1 = full NLD weighting)

As with the synthetic test signal, the TCD performs properly with the two real-world music signals. The virtual bass effect for the drum beat consists almost entirely of the NLD output due to the transient-rich signal content. The PV output is only mixed into the effect to enhance the resonances of the pitched drums after the initial transient attack. Similarly with the vocal sample, the NLD is only utilized for the initial onset of a note, while the PV handles the pitched components of the singing.

The tests in Fig. 7 and Fig. 8 confirm the appropriate functionality of the TCD for use within the hybrid virtual bass system.

VI. SUBJECTIVE EVALUATION

Subjective evaluation is required to adequately evaluate the hybrid system in comparison to pure NLD and PV systems since objective measurements cannot provide data for the virtual bass effect which occurs in the psycho-acoustical domain. The hybrid system should be less sensitive to signal content (i.e. music genre and tempo) giving consistent ratings across a wide variety of musical stimuli, while the NLD and PV should receive lower ratings for certain signal content due to their respective shortcomings.

A. The MUSHRA subjective testing method

The MUSHRA subjective testing method was developed to meet the needs of researchers in evaluating audio systems that cannot be classified objectively. The method, first developed

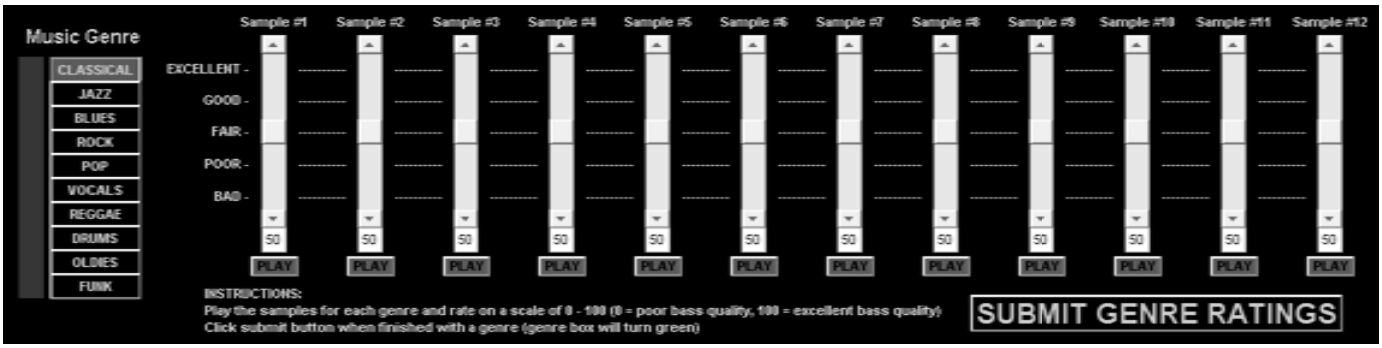

Figure 9. GUI used for subective virtual bass system evaluation

within the BBC, was passed by the International Telecommunications Union (ITU) as ITU-R Recommendation BS.1534 [10]. The method requires subjects to be presented with multiple stimuli (MUS) where they are all available for listening at any time during the test. This allows for subjects to directly compare the stimuli to determine the relative quality over the entire set.

Within each set of stimuli there must be a hidden reference and anchor (HRA). This should provide a benchmark for the system under test. In the case of virtual bass systems, the hidden reference is the unprocessed, full-range signal while the hidden anchor is the high-passed version of the original signal with no virtual bass. In a properly constructed subjective test, the reference signal should receive the highest subjective rating while the anchor should get the lowest.

The rating scale ranges from 0 (bad sound quality) to 100 (excellent sound quality). Descriptors are placed at 25 point intervals (poor, fair and good) to assist subjects to assign appropriate ratings.

B. Subjective evaluation software

A graphical user interface (GUI) was developed in MATLAB to meet the MUSHRA requirements (Fig. 9). The GUI presents subjects with ten stimuli, each containing different virtual bass enhancements applied to the same five second music sample. The ten virtual bass samples are randomly placed within the sample space along with the unprocessed signal (reference) and the high-passed signal (anchor).

Subjects proceed through ten different genres of music (in any order) and rate the low-frequency quality of each sample. This allows for the virtual bass systems to be rated across a wide range of stimuli where certain ones will favor NLD or PV processing.

C. Test procedure

Tests were carried out within a quiet, isolated listening room where subjects were left alone to complete the test with no time constraints. Subjects listened to the stimuli over a set of Beyerdynamic DT-770 headphones driven by a Sound Devices USBPre, connected directly to a PC. Sound levels could be adjusted to meet individual listening preferences.

While there was no time limit imposed on the subjects, it was recommended they spend approximately 3 – 5 minutes on each genre, resulting in a total test time of 30 – 50 minutes. Subjects were free to listen to stimuli as many times as necessary and could complete the genres in any order.

D. Test results

The tests included twenty-one subjects ranging in age from 23 to 63 (fifteen males and six females). Although subjects found the test demanding in terms of duration, they all completed the required tasks successfully, expressing overall enjoyment of the listening test.

The test results are presented in Fig. 10 where:

R = Reference (unprocessed signal)

A = Anchor (high-passed signal)

N = NLD system average

P = PV system average

H = Hybrid system average

with the results separated into their respective genres.

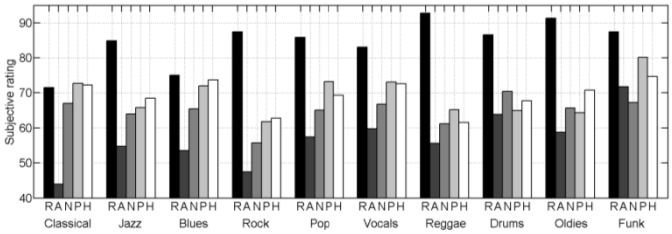

Figure 10. Average subjective ratings for virtual bass systems (R = reference, A = anchor, N = NLD, P = PV, H = hybrid)

The unprocessed (reference) sample achieved the highest ratings for all genres with the exception of classical. A suggestion for the lower classical rating is that symphonic instruments produce harmonically-rich notes with weak fundamental components. While perceived pitch remains centered at the fundamental frequency, the higher harmonics are more heavily weighted so that adding the virtual bass effect may be perceived similarly to the unprocessed sample.

In all cases except funk, the high-passed (anchor) sample received the lowest ratings, although in a number of cases the

anchor ratings were very close to the virtual bass ratings. These cases can be attributed to the virtual bass system under question introducing a noticeably artificial quality to the sample.

The hybrid and PV systems outperformed the NLD for all genres except drums and oldies. These two genres are very transient-rich, thus lending themselves nicely to the NLD approach. In both these cases, the hybrid approach outperformed the PV system, indicating that the TCD performed as expected, giving heavy weighting to the NLD output.

Although the hybrid virtual bass system did not receive the highest rating for all genres, it never received the lowest. In most cases where it was not the best rated the hybrid system rated closely below the PV system or split the difference between the PV and NLD systems. This indicates that the hybrid system is generally less sensitive to input material and can give consistent performance when incorporated into an audio system.

VII. CONCLUSIONS

A novel approach to the virtual bass for bandwidth extension of loudspeakers or bass boost has been described which utilizes a hybrid approach between nonlinear device (NLD) and phase vocoder (PV) based virtual bass systems. This hybrid system favors the NLD approach when presented a signal with high transient content, but relies on the PV method for steadier, pitched signals. The mixing function between the two devices is handled by a transient content detector (TCD) which operates in the frequency domain using constant-Q transforms (CQT) due to their increased low-frequency resolution capabilities over conventional time-frequency transforms, such as the discrete Fourier transform (DFT).

All three virtual bass systems (NLD, PV and hybrid) were evaluated subjectively to judge the perceived performance of each. Results indicate that while the hybrid approach is not always the highest rated virtual bass system, it is less sensitive to signal content fluctuations and showed no weaknesses for any of the musical genres included in the test.

Based on the subjects' feedback on the test, future tests will likely contain only five samples per genre (reference, anchor, NLD, PV and hybrid) as the extra variations for each virtual bass method caused the test to be tedious and time-consuming. These future tests would benefit from a shortened time frame as this would minimize any chance of listener fatigue, which can produce unreliable and unrepeatable results.

Overall, the hybrid virtual bass method proposed in this paper has performed well in comparison to the conventional NLD and PV techniques. This method can be applied to any application where more bass perception is necessary without concern for performance over a wide variety of signal content.

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] 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.

[3] Larsen, E; R.M. Aarts. "Reproducing low-pitched signals through small loudspeakers." Journal of the Audio Engineering Society, Volume 50, Number 3, pp. 147-164, March 2002.

[4] Daniel, B.T.; C. Martin. "The effect of the MaxxBass phychoacoustic bass enhancement system on loudspeaker design." 106th Convention of the Audio Engineering Society. Munich, Germany. May 1999.

[5] 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.

[6] Robel, A. "A new approach to transient processing in the phase vocoder." Proc. 6th Int. Conference on Digital Audio Effects. London, UK. September 2003.

[7] Nagel, F.; S. Disch; N. Rettelback. "A phase vocoder driven bandwidth extension method with novel transient handling for audio codecs." 127th Conference of the Audio Engineering Society. Munich, Germany. May 2009.

[8] Laroche, J.; M. Dolson. "New phase-vocoder techniques for real-time pitch shifting, chorusing, harmonizing, and other exotic audio modifications." Journal of the Audio Engineering Society, Volume 47, Number 11, pp. 928-936, November 1999.

[9] Brown, J.C. "Calculation of a constant Q spectral transform." Journal of the Acoustical Society of America, Volume 89, Number 1, pp. 425-434, January 1991.

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