Audio Bandwidth Extension

Somesh Ganesh
Georgia Tech Center for Music Technology
Georgia Institute of Technology
Atlanta, Georgia 30318
Email: someshg94@gatech.edu

***Abstract*—Audio bandwidth extension is a technique used to improve the perceived quality of band limited audio generated using various audio codecs. This paper proposes a few methods for audio bandwidth extension and evaluates them by performing listening tests. A comparison between half wave rectification and full wave rectification, along with the use of sub band filtering is done. The results obtained show that half wave rectification using sub band filtering is the optimal technique amongst the tested algorithms.**

I. INTRODUCTION

There is a huge set of band-limited audio in today's world. This is mainly due to the following reasons. Digitization of signals requires analog signals to be sampled. When the sampling rate in the analog to digital converter is not as high as desired, aliasing occurs. To avoid this, anti-aliasing filters are implemented as a pre-processing step. These filters are basically low pass filters which limit the bandwidth of the signal. Another reason for the vast number of band-limited audio is the algorithms implemented in audio codecs. To increase storage, audio codecs take advantage of the fact that humans are less sensitive to high frequency components as compared to low frequency components and compromise on the high frequency components of an audio file. This leads to an increase in storage and can significantly improve bit rate as well.

Band limiting reduces the quality of the audio. Hence, there is a need to develop an audio bandwidth extension algorithm to retrieve the original audio quality.

II. BACKGROUND

Let us begin by categorizing the related work into the two main categories, blind bandwidth extension and non-blind bandwidth extension.

A. Non-blind bandwidth extension

Non-blind bandwidth extension refers to the process of reconstructing the missing frequency spectrum using special encoding and decoding techniques. This involves including some of the timing or frequency information, noise levels or any other relevant information regarding the missing frequency components in the encoded signal along with the low frequency components. One of the most famous non-blind bandwidth extension algorithms is *Spectral Band Replication* (SBR) [1]. This method has its own encoder and decoder which works on the assumption that the low frequency and high frequency components have a strong correlation. It also accounts for signals having lower correlation by using techniques like inverse filtering, adaptive noise addition and sinusoidal regeneration. [2] introduces an algorithm using a *Fractal Self Similarity Model* (FSSM) for the *Modified Discrete Cosine Transform* (MDCT) representation of the audio signal. This method was shown to work on a variety of audio signals with detailed reconstruction of the missing frequency spectrum. [3] introduces *Accurate Spectral Replacement* (ASR), which reconstructs the tonal components and harmonic structures in the missing frequency spectrum. [4] uses the concepts in [2] and [3] and applies them to the high frequency resolution representation of the signal such as MDCT. It also includes "*Multi Band Temporal Amplitude Coding*" (MBTAC) which is used for the temporal shaping of the reconstructed high frequency components. The problem with these methods is that the process requires additional channel resources to transmit information about the missing frequency spectrum. This is not desirable when there is a constraint on the bit rate and storage.

B. Blind bandwidth extension

Blind bandwidth extension refers to the process of reconstruction without any prior information regarding the missing frequency spectrum. [5] propose an algorithm for blind bandwidth extension using half-wave rectification. This method involves half-wave rectifying the highest octave in the band limited signal to produce the high frequency spectrum. This new spectrum is scaled by a gain factor and added back to the delayed input signal. [6] looks to optimize this process for real time application by suggesting a different method for filtering the highest octave. This method also includes a bandwidth detection module as a preprocessing step and an adaptive gain as a post processing step. [7] and [8] suggest the use of linear extrapolation to find the envelope of and reconstruct the high frequency spectrum. The algorithm in [9] is based on *Phase Space Reconstruction* (PSR). Here, PSR is used to convert the low frequency MDCT coefficients of wideband audio into a multi-dimensional space. The high frequency spectrum is adjusted according to the listener's perception. This incorporates both linear and nonlinear prediction. [10] proposes an algorithm based on the chaotic prediction theory and suggests the generation of high frequency information based on the principles of audio production and the perception of the human ear.

III. Algorithm

The algorithm used in this paper is a blind bandwidth extension algorithm. This is because there is already a large set of band limited audio without the original audio. A non-blind extension would not help much in such a situation.

Let us first briefly look at the implemented system in Figure 1.

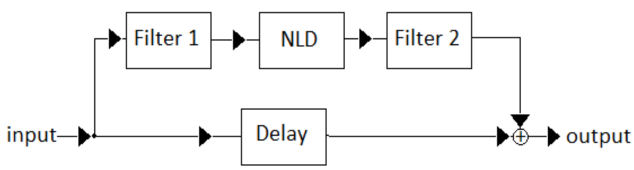

Figure 1. Algorithm overview

A. Filter 1

This section comprises the filtering operation done on the band limited audio. In this process, the highest octave in the signal is extracted. Assume the highest frequency component present in the signal is f_{hf}. Now, assume that the sampling frequency of the signal $f_s \geq 4*f_{hf}$, which can be done by upsampling as a preprocessing step. The highest octave present in the signal will be between $f_{hf}/2 - f_{hf}$. This filtered signal is used to generate the missing high frequency spectrum in the next block using a non-linear device. The non-linear device produces intermodulation distortion (defined in the next section). To study the effect of this distortion on the quality of the system, we have implemented two different versions of this block. The first one is a simple IIR (infinite impulse response) bandpass filter which passes frequencies in the range of $f_{hf}/2$ and f_{hf}. Intermodulation distortion increases with the amount of frequency components involved. Hence, the second version is a combination of two IIR bandpass filters where the first one filters out one half of the highest octave in the signal while the other one filters out the other half.

B. Non-Linear Device

This section consists of the non-linear device which generates the high frequency spectrum using the filtered signal from the previous block. It is the section where the higher harmonics of the signal are generated. Here, we have made use of two non-linear devices. One is a half wave rectifier and the other is a full wave rectifier. We have chosen rectification for this section as it is a homogeneous process. The output of a rectifier mainly consists of the second harmonic of the input frequency with a decay of 12 dB per octave. The spectrum of a half wave rectified output consists of the original input frequency as well as all of its harmonics. On the other hand, the spectrum of a full wave rectified output consists of only the even harmonics of the input frequency. We are interested in generating the next harmonic, i.e., the spectrum between $f_{hf} - 2*f_{hf}$. Now, let us define the term intermodulation distortion (IMD). When a signal consisting of two or more frequency components goes through a non-linear operation, the output not only contains these frequency components but also several undesirable components which are basically the sum and difference of the input components. This distortion in the output is known as intermodulation distortion. In this paper, we study the effect of intermodulation distortion of half wave and full wave rectification on the subjective quality of the reconstructed audio signal.

C. Filter 2

This section follows the non-linear device block. We are only interested in the high frequency spectrum generated in the previous block. Following the rectification process in the previous block, the high frequency components will mainly contain the second harmonics of the input frequency components. Due to intermodulation distortion, there will be undesirable components present below f_{hf} and above $2*f_{hf}$. These components need to be eliminated. Hence, the interested signal will lie between f_{hf} and $2*f_{hf}$. The signal between these boundaries will be filtered out in this stage. Similar to the Filter 1 block, we will have two versions of this block. The first version will contain a simple bandpass IIR which will filter out the signal between f_{hf} and $2*f_{hf}$. The second version will contain two separate IIR bandpass filters. These filters will be used when the second version of Filter 1 is used. The first filter will filter out the second harmonics of the input frequencies from the first filter in the second version of the Filter 1 block. The second filter will filter out the second harmonics generated from the input frequencies from the second filter in the second version of the Filter 1 block. The output from these two filters can then be added to give us the complete reconstructed high frequency spectrum.

Finally, the generated high frequency spectrum is added to the input which is delayed by an amount equivalent to the time taken to generate the spectrum. Hence, we will finally have the bandwidth extended signal.

IV. Evaluation

A. Experiment Methodology

For the evaluation, listening tests were performed to understand the subjective quality of the bandwidth extended audio. The dataset for the experiment was generated using a simple IIR low pass filter. Five songs were chosen from different genres – electronic, rock, electro house, acoustic and world music. All the files had a sampling frequency of 44.1 KHz and were band limited to 7 KHz. Each file was 20 seconds long.

For the listening test, a *MUSHRA* [11] like test was conducted. The participants were first made to hear the original audio file (reference) followed by the band limited audio file. After this, they were made to hear five files in random order which included a hidden reference. The five files included files which were bandwidth extended using the algorithms mentioned above –

- Half wave rectification
- Half wave rectification using sub band filtering
- Full wave rectification
- Full wave rectification using sub band filtering

They then rated the files on a scale from 1 to 10, 1 being the least amount of perceived quality and 10 being the maximum amount of perceived quality. This was repeated for all five songs. The loudness for all these files were normalized to a RMS (root mean square) value of 0.7.

The spectrograms for all 5 files of one of the songs as well as the band limited signal has been shown in Figure 2 below.

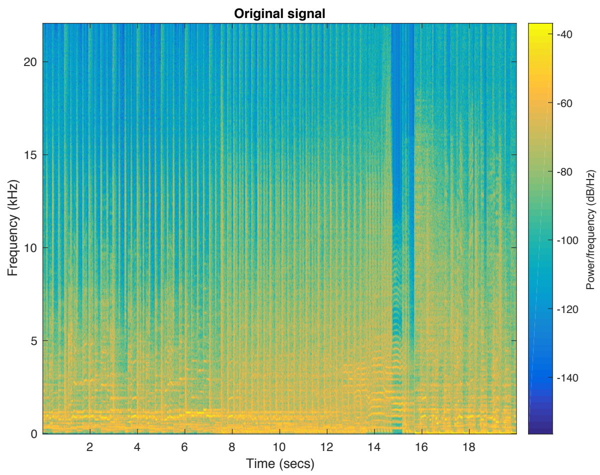

Figure 2a. Original signal (Reference)

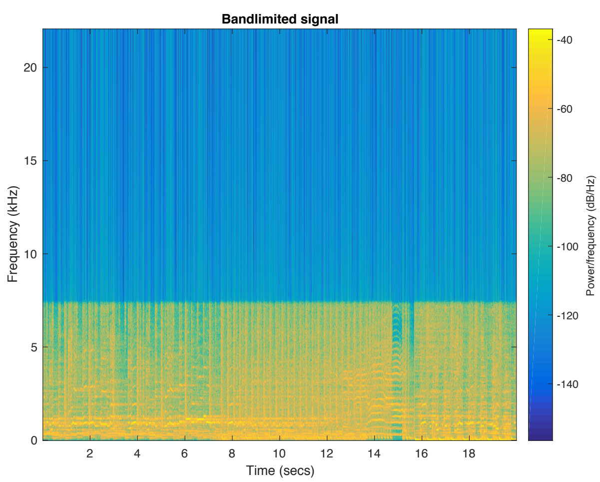

Figure 2b. Band limited signal

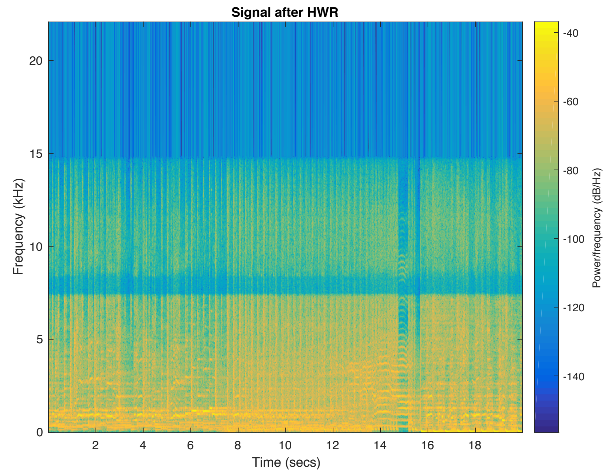

Figure 2c. Reconstructed signal using Half wave rectification

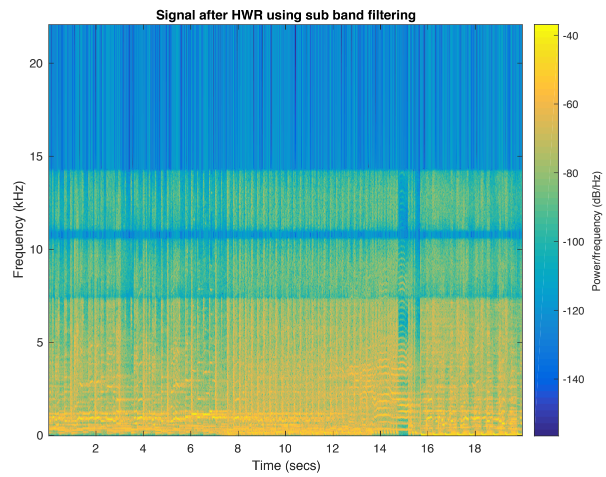

Figure 2d. Reconstructed signal using Half wave rectification and sub band filtering

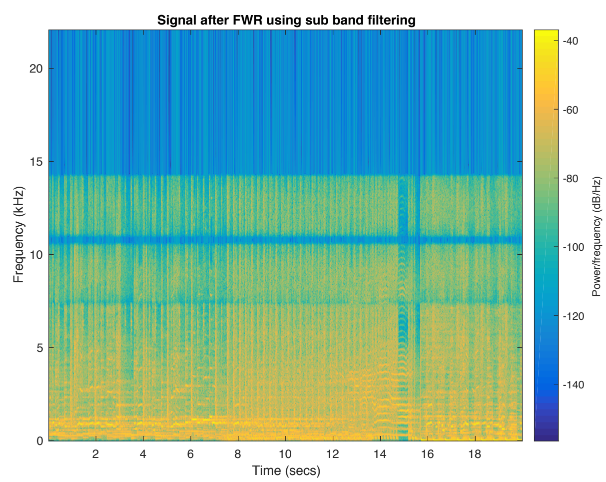

Figure 2e. Reconstructed signal using Full wave rectification

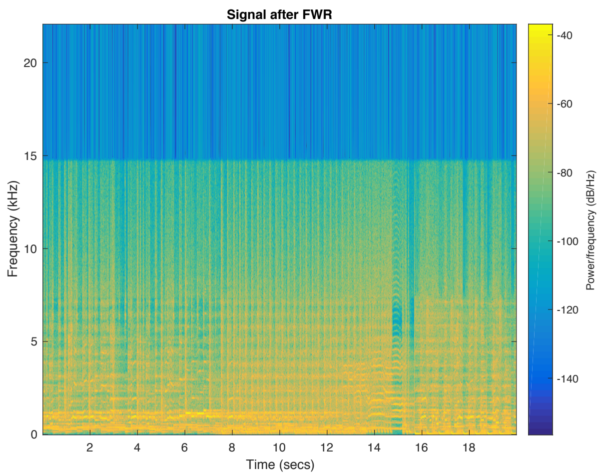

Figure 2f. Reconstructed signal using Full wave rectification and sub band filtering

B. Results

The test had 6 participants. The results obtained were recorded and the means of the ratings were calculated. These means are tabulated in Table 1.

	Reference	Half Wave Rect.	Half Wave Rect. (sub band filtering)	Full Wave Rect.	Full Wave Rect. (sub band filtering)
Song 1 – electronic	8.83	4.83	4.83	3.5	5.67
Song 2 – acoustic	9.5	6.33	5.67	3	5.67
Song 3 – world	8.67	5.67	4.5	4.5	5.67
Song 4 – electro house	8.5	5.5	5.83	3.16	4.67
Song 5 – rock	8.16	3.33	4.67	1.83	4.16

Table 1. Means of ratings.

From the results, we can see that the original audio file (reference) receives the highest rating. This tells us that the ratings given by the participants are genuine and can be considered.

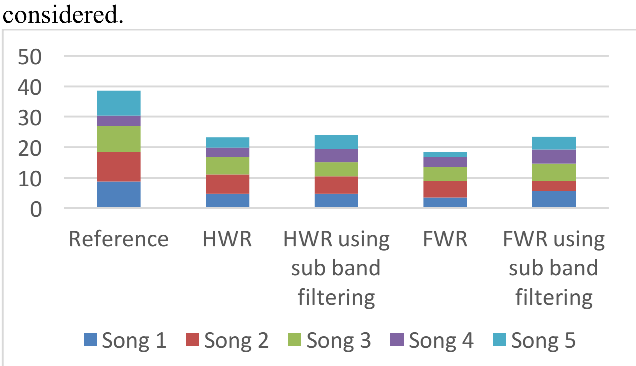

Figure 3 – Bar graph of means of ratings.

From the bar graph in Figure 3, we can see that the original audio files have much better quality than the bandwidth extended files. We can also observe that the use of sub band filtering in the algorithm had a positive effect on the subjective quality of the audio files. The files extended using simple full wave rectification were found to have the least amount of perceived quality amongst all the files.

V. Discussion

First, let us discuss the results obtained using the algorithm and experiment presented in this paper.

From the results, we can conclude that the use of sub band filtering decreases the amount of intermodulation distortion in the reconstructed output and hence has a higher subjective quality. We can also conclude that half wave rectification is a better approach as opposed to full wave rectification. From the rating received, we can rank the algorithms in terms of the amount of perceived quality as

1) Half wave rectification using sub band filtering
2) Full wave rectification using sub band filtering
3) Half wave rectification
4) Full wave rectification

These results validate the method suggested by [6] where the use of sub band filtering is expected to have a higher perceived quality but is not compared to the general filtering technique used in [5]. This paper also compares the effect of half wave rectification vs. full wave rectification on the perceived quality.

Now, let us discuss the shortcomings of the algorithm and experiment presented in this paper.

The algorithm proposed here can reconstruct the frequency spectrum up to the maximum limit (fs/2) only if the highest frequency present in the bandlimited signal $f_{hf} = f_s/4$. If $f_{hf} < f_s/4$, as shown in Figure 2, we cannot reconstruct the high frequency spectrum up to the maximum range.

The experiment conducted had very few participants and a small dataset.

VI. Future Work

This paper can be used as a basic frame work for future work in audio bandwidth extension. The following can be considered as improvements and extensions for the algorithms and experiment presented in this paper

- Reconstruction of the high frequency spectrum can be made independent of the highest frequency present in the band limited audio. This will allow suitable reconstruction even if the highest frequency present in the signal is very low, i.e., in the range of 5 – 7 KHz.
- Since only two different rectifiers have been used in the non-linear device block, further investigation can be done to test different non-linear devices such as an integrator.
- This whole system can be optimized further to work much better in real time.

VII. CONCLUSION

This paper has compared different methods of audio bandwidth extension and evaluated the results by performing listening tests. The conclusion from the experiment conducted is that half wave rectification performs better than full wave rectification as a non-linear device and that the use of sub band filtering improves the amount of the perceived quality in the bandwidth extended signal.

VIII. REFERENCES

[1] P. Ekstrand, "Bandwidth extension of audio signals by spectral band replication," in Proceedings of the 1st IEEE Benelux Workshop on Model Based Processing and Coding of Audio (MPCA '02, 2002.)

[2] Deepen Sinha, Anibal Ferreira, and, Deep Sen "A Fractal Self-Similarity Model for the Spectral Representation of Audio Signals," 118th Convention of the Audio Engineering Society, May 2005, Paper 6467.

[3] Anibal J. S. Ferreira and Deepen Sinha, "Accurate Spectral Replacement," 118th Convention of the Audio Engineering Society, May 2005, Paper 6383.

[4] H. E. V., A. J. S. Ferreira, and D. Sinha, "A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)," presented at the Audio Engineering Society Convention 120, 2006.

[5] E. Larsen, R. M. Aarts, and M. Danessis, "Efficient High-frequency Bandwidth Extension of Music and Speech," presented at the Audio Engineering Society Convention 112, 2002.

[6] M. Arora, J. Lee, and S. Park, "High Quality Blind Bandwidth Extension of Audio for Portable Player Applications," presented at the Audio Engineering Society Convention 120, 2006.

[7] C.-M. Liu, W.-C. Lee, and H.-W. Hsu, "High Frequency Reconstruction for Band-Limited Audio Signals," in Proceedings of the 6th International Conference on Digital Audio Effects (DAFX-03), 2003.

[8] "AES E-Library » High Frequency Reconstruction by Linear Extrapolation." [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12426.

[9] C.-C. Bao, X. Liu, Y.-T. Sha, and X.-T. Zhang, "A blind bandwidth extension method for audio signals based on phase space reconstruction," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2014, no. 1, p. 1, 2014.

[10] Y. T Sha, C. C Bao, M. S Jia, and X. Liu, "High frequency reconstruction of audio signal based on chaotic prediction theory," in 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, pp. 381–384.

[11] The MUSHRA audio subjective test method - BBC R&D, 2002.