Audio Engineering Society
Convention Paper 6761
Presented at the 120th Convention
2006 May 20–23 Paris, France



# High Quality Blind Bandwidth Extension of Audio for Portable Player Applications

Manish Arora, Joonhyun Lee, and Sangil Park

DM R&D Center, Samsung Electronics Co. Ltd, 416, Maetan3-dong, Suwon-si, 443742-South Korea
manish.arora@samsung.com, johny8.lee@samsung.com, sipark@samsung.com

## ABSTRACT

Bandwidth limitation in lossy audio coding schemes significantly reduces the perceived quality. High frequency bandwidth extension schemes have been proposed but are difficult to implement in applications where they are needed most, in portable audio devices with severe complexity constraints. The following work describes a high quality blind bandwidth extension method proposing efficient initial audio band width detection, band based nonlinear processing and simple regenerated spectral envelop shaping enhancements. Objective and Subjective measurements of the processed signal have yielded significant quality improvements with very low complexity requirements allowing easy implementation on a wide variety of portable player platforms.

## 1. INTRODUCTION

Most perceptual audio codecs bandwidth limit the input signals in order to use the available bits for psychoacoustically relevant low frequency signals. The bandwidth limitation becomes severe at very low bit rates e.g. 128kbps mp3 encoded audio is band width limited to about 15 kHz and 64kbps mp3 to about 8 kHz. Figure 1 shows an example of a 96kbps mp3 encoded-decoded signal spectrum, the signal spectrum chopped beyond 11 kHz. These missing high frequency signals cause audio signals to sound dull and at times muffled. The goal of this work is to improve audio quality in such cases by artificially creating high frequencies from low frequency information.

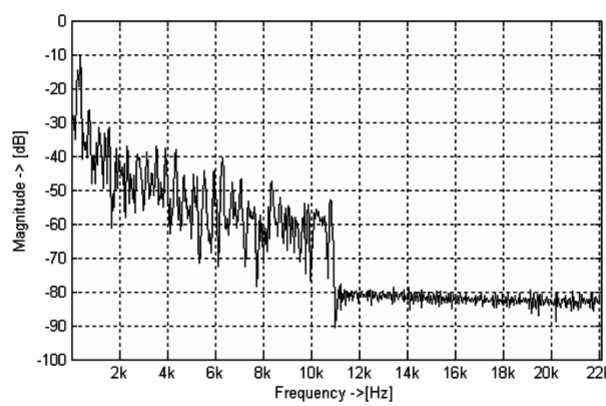


Figure 1: Bandwidth limitation in compressed audio.

Recently proposed encoder side modifications [1] solve the bandwidth limitation issues but there is an enormous amount of content available which suffers from the above mentioned quality issues. Hence from a portable audio player designer's perspective it is still essential to include solutions which serve the existing content. Post processing methods after decoding have also been proposed recently such as the spectral extrapolation method [2], which converts the signal to frequency domain and performs linear extrapolation on the magnitude spectrum. Unfortunately the method is computationally too expensive to be easily implemented on portable audio players. Instead, time domain methods are more suited for easy real time implementation in portable devices where computational complexity is very limited.

Proposed time domain methods [3] involve filtering out some part of decoded signal and then using non linearity to generate harmonics. The harmonics created are post filtered and added back to the original signal to obtain the enhanced signal. The effects of various non linear operations on audio signals also have already been well studied [4]. One of the remaining issues of existing time domain approach is that the bandwidth of the signal is unknown. Hence the starting point of missing frequencies is not known and perhaps can only be guessed from bit rate of the encoded signal or from the decoding information. At times it might be unpractical to receive information from the decoding system or the bandwidth and thus the starting point of the missing frequencies might be varying with time as in a variable bit rate system. So it becomes necessary to detect the bandwidth of the decoded signal in real time. Existing approaches also suffer from excessive intermodulation noise problems. While nonlinear processing generates harmonics a lot of intermodulation noise is also generated. Ways to implement nonlinear processing with reduced intermodulation distortion would improve the performance. Lastly suitable low complexity reconstructed spectrum shaping methods are required to ensure that the reconstructed spectrum has continuity and smoothness with the decoded signal spectrum.

In this paper we propose enhancements to the mentioned problems of existing time domain high frequency reconstruction methods. The target complexity of our proposed algorithm is about 20MIPS on ARM processor allowing for easy and widespread deployment on various portable player models. Section 2 describes the details of the proposed algorithm involving low complexity bandwidth detection, subband based nonlinear processing and recreated subband gain calculation for spectral shaping. Section 3 shows the performance of our algorithm and analysis of results followed by conclusions and references in sections 4 and 5.

## 2. ALGORITHM DESCRIPTION

The high frequency reconstruction algorithm receives decoded data as input from any lossy audio decoder and recreates high frequencies blindly i.e. by using only the decoded audio signal and nothing else. Since the bandwidth of the input signal is unknown it is first estimated by using a real time bandwidth detection process. After detecting the highest frequency present in the signal at any given time, the decoded signal is divided into subbands up to half the detected highest frequency of the signal. Each subband signal are then individually passed through non linearity to generate harmonics. The generated harmonics are gain scaled to achieve spectral envelop shaping and added back to the original signal. Figure 2 shows the overall algorithm block diagram.

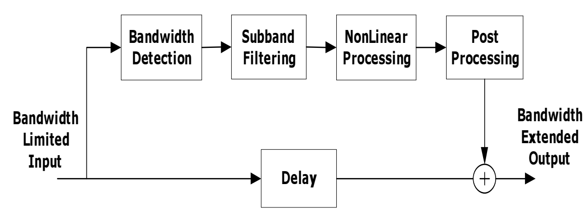


Figure 2: High frequency reconstruction algorithm.

### 2.1. Bandwidth Detection

Bandwidth detection involves analyzing the signal in real time e.g. every 20msec's to find the highest frequency present in the signal. A typical range of bandwidth of interest is from 7 kHz to 16 kHz. Signals with bandwidth below 7 kHz might mean that the original audio signal does not contain high frequencies. Bandwidth beyond 16 kHz means that there are significant amount of high frequencies present and reconstruction is not required. A possible solution for bandwidth detection might be to perform an FFT on the decoded audio data. Using the frequency response it would be simple to detect the highest frequency with any significant energy. But an FFT would increase the complexity of the reconstruction algorithm and application to various platforms would be difficult. Instead of using a FFT we could have a bank of bandpass IIR filters followed by energy calculation of filtered results. For 1 kHz accuracy in detection we need to have 8 bandpass filters in our frequency range of interest. Even if we use 2nd order IIR filters the complexity requirements are very high. Instead a very low complexity method which involves a single multiple accumulate per bank is used. We generated white noise signal offhand and windowed it with a short time hamming window. This signal is bandpass filtered using filters with very sharp cutoff frequencies offline to create bandpass filtered noise signals. We created 8 such signals from 7 kHz to 16 kHz each with 1 kHz bandwidth from the windowed white noise signals. Windows of length 128 samples at 44.1 kHz sampling was found to be sufficient. These bandpass filtered noise signals are stored and the input signal is cross correlated with each of the noise signals during the bandwidth detection step. The cross correlated results are compared in order to find the highest frequency with significant energy which allows us to determine the bandwidth of the signal. Hence we are able to detect the bandwidth of the signal within 1 kHz accuracy at a complexity of a few multiply accumulate operations per input sample. Figure 3 shows the bandwidth detection procedure.

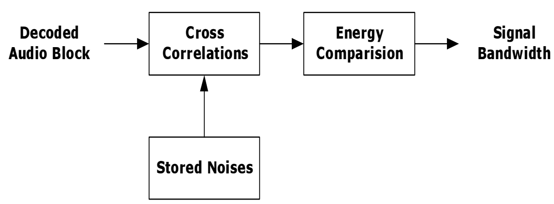


Figure 3: Bandwidth detection algorithm.

### 2.2. Subband filtering

Bandwidth detection step is followed by subband filtering in which the input signal is divided into multiple bands before the nonlinear processing step. It is well known that any nonlinear operation produces harmonics of the input frequency along with intermodulation noise. The more the frequency contents are present in the signal before the nonlinear process the more the intermodulation noise. Hence if we bandpass filter the input signal into multiple bands and apply nonlinear processing individually on each subband the intermodulation noise would be reduced. The higher the number of subbands the lesser will be the intermodulation noise. In our implementation we divided the signal into 2 subbands. Although higher number of subbands would have been desirable, the complexity of implementation would have been significantly increased. If B is the bandwidth detected then the 1st band extends from 0.5xB – 0.75xB and 2nd band from 0.75xB – B. The 1st and 2nd subbands are called Band1 and Band2. Since the detected bandwidth is approximated to be within a set of fixed values, the filter coefficients used for subband filtering are pre-calculated and stored. 4th order IIR filters were found to be sufficient for the purpose. The filters were implemented using the load store reduction method suitable for RISC processors as described in [5]. Linear phase IIR filtering methods [6] were also implemented but were later found to not provide sufficiently high quality gains corresponding to the complexity increase.

### 2.3. Nonlinear Processing

The subband filtered data is then passed through a non linear process to detect harmonics. Any non linear process which generates second harmonic can be used. Full wave rectification is a suitable non linear process. Along with very low complexity of implementation it also exhibits harmonic signal amplitude linearity [4]. The full wave rectification process is followed by bandpass filtering in order to filter out frequencies outside the range of interest generated due to intermodulation and aliasing. Band1 harmonics are post filtered from B – 1.5xB and 2nd subband harmonics from 1.5xB – 2.0xB. The result of post filtering of Band1 data is called Band3 and that of Band2 is called Band4. Figure 4 shows bands Band1, Band2, Band3 and Band4. If B is detected greater than 9 kHz, the maximum recreated frequency is limited to 18 kHz.

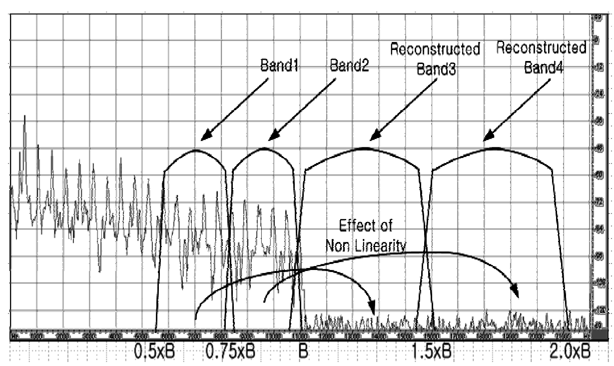


Figure 4: Input and reconstructed subbands.

### 2.4. Post Processing

The goal of the post processing step is to modify the energy of the recreated bands so as to have them match the original frequency content. A simple algorithm that attempts to maintain the continuity of the spectral envelope by modifying energies of Band3 and Band4 has been developed. After the nonlinear processing and post filtering step the energy in bands Band1, Band2, Band3 and Band4 are evaluated on a short time basis. The energy of the original signal's subband Band1 is called $E_1$, that of Band2 is called $E_2$ and those of reconstructed Bands 3 and 4 are $E_3$ and $E_4$ respectively. Assume that $E_{3Target}$ and $E_{4Target}$ are the desired energies of Band3 and Band4 respectively which would give a smooth reconstructed spectral envelope. From our experiments we found that the desired value of $E_{3Target}$ is approximated by equation (1).

$$E_{3Target} \approx \frac{E_2 \times E_2}{E_1} \qquad (1)$$

Using the expression from equation (1) the gain for Band3 $G_3$ can be calculated as,

$$G_3 \approx \frac{E_{3Target}}{E_3} \qquad (2)$$

where $E_3$ is the energy of the reconstructed Band3 after the non linear processing. Using calculations similar to equations (1) and (2) the target energy $E_{4Target}$ and gains required for reconstructed Band4 $G_4$ would be,

$$E_{4Target} \approx \frac{E_{3Target} \times E_{3Target}}{E_2} \qquad (3)$$

$$G_4 \approx \frac{E_{4Target}}{E_4} \qquad (4)$$

The reconstructed subbands Band3 and Band4 are then gain modified using $G_3$ and $G_4$ from equations (3), (4) and added back to the original signal to get the bandwidth extended output signal.

## 3. RESULTS AND ANALYSIS

In our work we measured the quality improvements of bandwidth extension using the perceptual evaluation of audio quality (PEAQ) system [7]. PEAQ uses a perceptual model measuring nonlinear distortion, linear distortion, harmonic structure, distance to masked threshold and changes in modulation to evaluate differences in audio quality between two tracks. The output of the PEAQ system is an objective score called as the objective difference grade (ODG). ODG values vary from 0 to -4 with 0 being imperceptible loss in quality and -4 being a very annoying degradation in quality. ODG score were first calculated for the decoded signal compared to the original signal. Next the decoded signal was bandwidth extended and ODG score was again evaluated with respect the original signal. By comparing the ODG scores before and after bandwidth extension, the gain due to the process was calculated. A total of 40 popular audio tracks were prepared each of length 60 seconds. The 40 tracks constituted 10 tracks each of the music genres classic, pop, rock and jazz. Each of the 40 tracks were mp3 encoded using 64kbps, 96kbps and 128kbps bit rates.

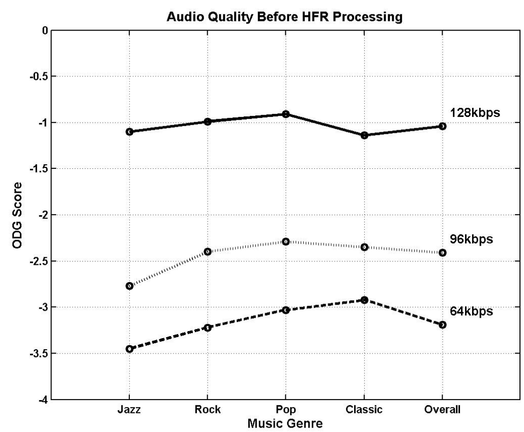


Figure 5: Audio quality before processing.

Figures 5 show the average ODG score obtained for 64bps, 96kbps and 128kbps decoded signals on each music genre basis before application of bandwidth extension. Two significant observations can be made from the plots. Firstly, at low bit rates Jazz and Rock music exhibited low quality. This is possibly because Jazz and Rock include more high frequency components than other genres, which are restricted severely at low bit rates leading to lower quality. Secondly, quality increases were seen with the increase of bit rates but a more significant improvement was observed when the bit rate increased from 96 to 128kbps for Jazz and Rock music. This is possibly attributed to the fact that mp3 encoded audio at 128 kbps has higher bandwidth capacity, hence can much better encode Jazz and Rock music.

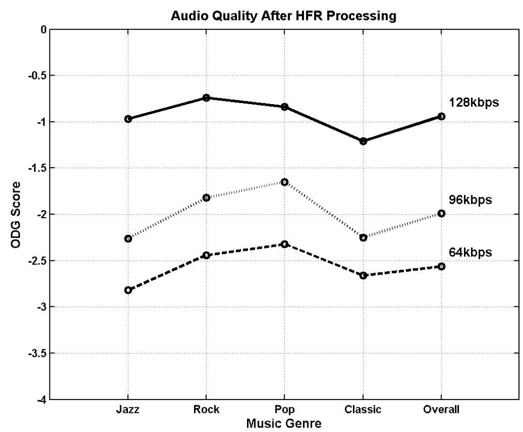


Figure 6: Audio quality after bandwidth extension.

Figure 6 shows the average ODG score obtained for 64bps, 96kbps and 128kbps decoded signals on each music genre basis after application of the proposed algorithm. Very high improvements in quality were observed for 64kbps signals regardless of music genre with the quality of 64kbps processed signal approaching that of 96kbps unprocessed signals. Significant enhancement was also observed for 96kbps signals with an average increase of 0.44 in ODG score. 128kbps signals show small improvements in quality. This may be because our algorithm cannot generate high frequencies to match the energy and spectral envelope of the original content. In the case of Classical music at 128kbps a reduction in quality due to the application of our algorithm was observed. This is because at times our algorithm regenerates frequencies which were not present in the original classical music signals in which the signal bandwidth variations are more rapid than other music genres. As a general trend it was observed that the increases in quality of Jazz, Rock and Pop were much higher than that of Classical music. This is perhaps because of the use of percussion and synthetic instruments in Jazz, Rock and Pop music which are well simulated by harmonics and intermodulation noise generated by our non linear processing approach. Figure 7 summarizes the gains observed after applying bandwidth extension algorithm. Figure 8 shows two examples of reconstructed spectrum for audio signals strong in harmonic content and with wideband signal content. The original spectra are offset by 50dB.

Subjective listening tests were carried out using informal listening. Listeners were made to listen to low bit rate decoded and enhanced audio clips and asked to choose the preferred signal. For 64kbps and 96kbps signals users clearly showed a preference for bandwidth extended signals. Users were in most cases unable to distinguish bandwidth extended and decoded signals for 128kbps signals.

The real time implementation of the algorithm was done on the ARM 926EJ-S processor at a complexity of 19MIPS. A low MIPS version involving joint high frequency reconstruction involving the same techniques was also developed at a complexity of 14MIPS.

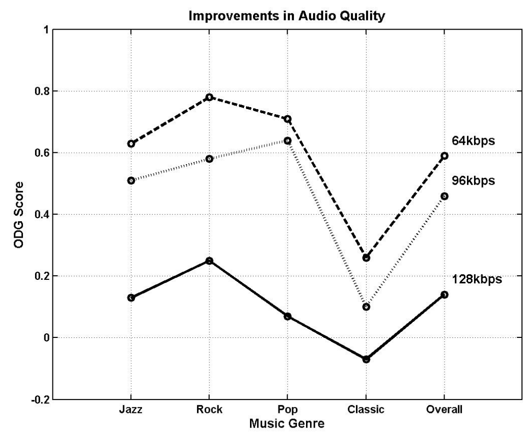


Figure 7: Enhancement in Audio quality after bandwidth extension algorithm application.

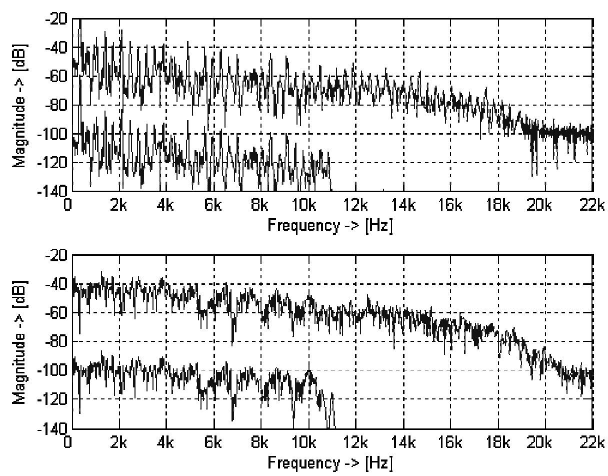


Figure 8: Bandwidth enhanced and original spectrums.

## 4. CONCLUSIONS

The following paper describes a technique to recreated high frequencies lost due to lossy audio coding at a very low complexity without any encoder changes. The proposed method involves fast bandwidth detection, splitting the signal into bands, applying non linearity on a band by band basis and performing spectral shaping by controlling the energy of harmonic bands. A real time implementation of the proposed algorithm needed a complexity of 19MIPS allowing for easy and widespread deployment. Objective and Subjective measurements of the quality of the algorithm have yielded significant improvements especially when applied to Jazz, Rock and Pop music signals encoded at bit rates below 128kbps.

Although next generation portable audio players are expected to have enough memory to carry high bit rate audio, the widespread prevalence of low bit rate audio content today warrants the deployment of the above mentioned work for all portable audio players.

## 5. REFERENCES

[1] M. Dietz, L. Liljeryd, K. Kjorling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," presented at the 112th AES convention, Munich, Germany, 2002, May 10-13.

[2] Chi-Min Liu, Wen-Chieh Lee and Han-Wen Hsu, "High frequency reconstruction by linear extrapolation," presented at the 115th AES convention, New York, USA, 2003, October 10-13.

[3] Erik Larsen, Ronald M. Aarts and Michael Danessis, "Efficient high frequency bandwidth extension of music and speech," presented at the 112th AES convention, Munich, Germany, 2002, May 10-13.

[4] Erik Larsen and Ronal M. Aarts, *Audio Bandwidth Extension.* John Wiley and Sons, 2004.

[5] Manish Arora, Suresh Babu P. V. and Vinay M. K., "RISC processor based speech codec implementation for emerging mobile multimedia messaging solutions," 14th international conference on digital signal processing, DSP 2002, Santorini, Greece, July 1-3, 2002.

[6] S. R. Powell and P. M. Chau., "A technique for realizing linear phase IIR filters," IEEE Transactions on Signal Processing, vol 39(11), pp. 2425-2435, 1991.

[7] ITU Radiocommunication Study Group 6, "Draft revision to recommendation ITU-R BS.1387 – Method for objective measurements of perceived audio quality".