High Frequency Enhancement of Sound Using Wavelet Transform

Won-Jung Yoon*, Kang-Kyu Lee* and Kyu-Sik Park*
* Division of Information and Computer Science, Dankook University, Seoul, 140-714, Korea
Tel : +82-2-709-2728 Fax : +82-2-796-2970 E-mail: helloril@dankook.ac.kr, fitz@dankook.ac.kr, kspark@dankook.ac.kr

Abstract:
This paper proposes new method for the enhancement of nonexistent high frequency spectral contents from low sample rate audio signal. For example, Due to the protocol constraint, the audio bandwidth of MP3 is restricted to 16Khz. Although band-restricted MP3 audio provide savings of storage space and network bandwidth, it suffers a major problem of a loss in high frequency fidelity such as localization, ambient information, and bright nature of audio. This paper provides a new mathematical analysis for the adaptive estimation of the high frequency contents based on the nature of the input low sample rate audio. Proposed method can be worked globally to any kind of audio such as speech and music that are restricted by sampling rate and bandwidth.

Keywords: Wavelet Transform, High Frequency Enhancement, MP3

1. INTRODUCTION

Internet provides easy and minimal cost way to distribute and download audio and it brought new paradigm of the internet-based audio disseminating. However, the internet distribution of the full CD quality audio poses a new set of problem because it requires to process large amount of digital data in a sample rate of 44.1Khz. Consequently, this calls for the low sample rate audio processing or the audio data compression technology such as MP3. We note that the audio bandwidth of MP3 is restricted to 16Khz due to the protocol constraint and it again can be treated as low sample rate audio. Although these low sample rate audios provide savings of storage space and network bandwidth, it suffers a major problem such as a loss of high frequency fidelity.

Traditionally the low frequency content of audio is treated more important in terms of signal energy, but the high frequency contents also have lots of information such as localization and ambient information, and bright nature of audio.

There have been a few approaches to solve this kind of problem in the literature. Some attempts have been made to extrapolate a wideband signal from its narrowband frequency components [1]-[4]. However, they were all limited to speech, instead of a general audio. In paper [5], they proposed a method for high frequency reconstruction for band-limited audio signals using the least squares method in frequency domain. In paper [6], Corey proposed a direct method for the compensation of high frequency that is not present in low sample rate audio. Their paper used a DWT (Discrete Wavelet Transform) analysis/ resynthesis method to compensate high frequency content in the wavelet domain. The estimation is based on the exponential sum of DWT coefficients expanded from the low sample rate audio and the compensation is then performed by the DWT synthesis. However this method is quite heuristic problem setup because the estimation is solely relied on the fact of exponential decay characteristics of DWT coefficients as described in paper [7].

The purpose of this paper is to solve the same set of problem for the enhancement of nonexistent high frequency spectral contents from the low sample rate audio. The proposed algorithm differs from the previous approach in paper [6] in terms of high frequency content estimation method. Our approach provides a mathematical analysis for the adaptive estimation of the high frequency contents based on the nature of the input low sample rate audio.

2. PROPOSED METHOD

In this section, a new approach for the estimation and compensation of high frequency spectral contents from the low sample rate audio is proposed. This new method focuses directly on the estimation of first level DWT high frequency coefficients and compensation by the DWT synthesis. First, the principle of this approach is explained and then an adaptive algorithm is presented for the spectral estimation and compensation.

2.1 Principle

For the given low sample rate audio $x(n)$, it can be considered as a first level DWT low frequency coefficients $c_0(n)$ in the DWT analysis procedure as in Fig. 1. Then the estimation problem for the high frequency spectral contents from $x(n)$ can be interpreted as to estimate a first level DWT high frequency coefficients $d_0(n)$. The high frequency compensated audio can then be reconstructed through the DWT synthesis procedure. This idea is described in Fig. 2.

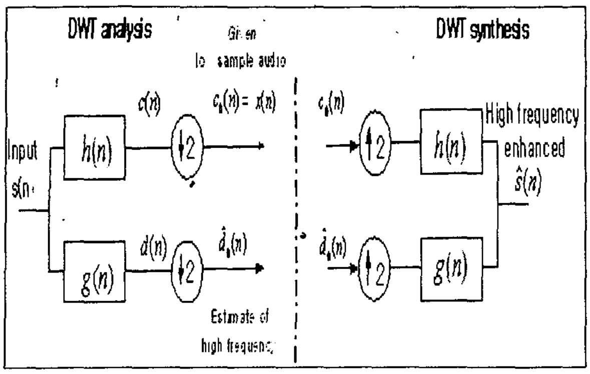


Fig. 1. DWT analysis and synthesis for high frequency spectral enhancement

Fig. 2 describes detailed DWT analysis and synthesis procedure to explain the estimation of missing high frequency $\hat{d}_0(n)$ for the given low sample rate audio $x(n) = c_0(n)$ and the reconstruction of high frequency enhanced audio $\hat{s}(n)$. In this figure, we note that $c(n), d(n)$ are the interpolated version of $c_0(n), \hat{d}_0(n)$ respectively. The high frequency estimate $\hat{d}_0(n)$ can be obtained simply from downsampling $d(n)$ by a factor of 2. By summarizing the above procedure, the proposed algorithm now can be implemented as following simple two steps. For the given low sample rate audio $x(n)$ and wavelet LPF and HPF $h(n), g(n)$, interpolate $x(n)$ by 2 and get $c(n)$. (1) Adaptively estimate $d(n)$ base on the observation of $c(n)$ and get DWT high frequency coefficients $\hat{d}_0(n)$ simply from downsampling $d(n)$ by a factor of 2.
(2) Reconstruct the final high frequency enhanced audio $\hat{s}(n)$ using the DWT synthesis procedure as in Fig. 2.

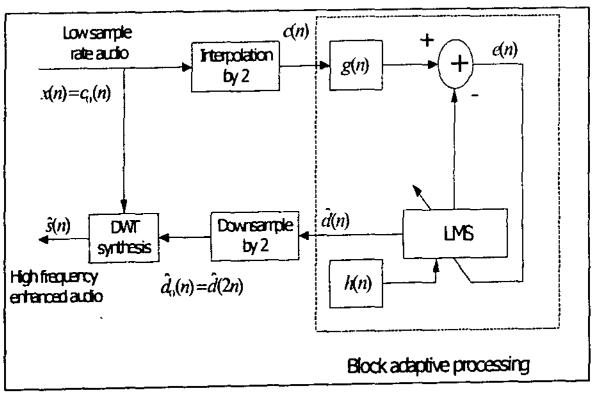


Fig. 2 LMS adaptive block diagram for the estimation of $\hat{d}(n)$ and reconstruction of $\hat{s}(n)$

2.2 Adaptive Estimation and Compensation Algorithm

We assume that the system is linear and time-invariant before the downsampling as in Fig. 1. Then we can have following relation in time n

$$s(n) * h(n) * g(n) = s(n) * g(n) * h(n)$$
$$e(n) = c(n) * g(n) - d(n) * h(n) = 0 \quad (1)$$

where $s(n)$ is an assumed input source signal, and $h(n)$, $g(n)$ are wavelet low-pass and high-pass impulse response, and $s(n) * h(n) = c(n)$ and $s(n) * g(n) = d(n)$.
With a reasonable interpolation method available, $c(n)$ then can be approximated from $c_0(n)$ as $c(n) = s(n) * h(n) \cong [c_0(n)]_{\text{intp\_by2}}$. Then Eq. (1) can be rewritten as

$$e(n) = [c_0(n)]_{\text{intp\_by2}} * g(n) - d(n) * h(n) = 0 \quad (2)$$

And the estimation problem can now be implemented as to find $\hat{d}(n)$, estimate of $d(n)$, that minimize error $e(n)$. For this kind of implementation, LMS (Least Mean Square) algorithm is often used in the literature [6]. With an error in Eq. (2), the LMS algorithm may be expressed as

$$\hat{d}(n+1) = \hat{d}(n) + 2\mu e(n) h(n) \quad (3)$$

where $c_0(n) = x(n)$, $\mu = u / L\sigma^2$, $0 < u < 1$ is the adaptation step size, L is an order of adaptive filter and σ^2 is energy of LPF $h(n)$.
After the estimation of $\hat{d}(n)$, we can obtain true high frequency counterpart of DWT coefficients by downsampling factor of 2 as $\hat{d}_0(n) = \hat{d}(2n)$. We then can reconstruct the high frequency enhanced audio $\hat{s}(n)$ by the DWT synthesis procedure with a complete set of DWT coefficients that compromise low sample rate audio $x(n)$ and the high frequency estimation $\hat{d}_0(n)$. This completes our analysis and synthesis procedure for the high frequency spectral enhancement of audio.
Now several discussions can be made for the proposed method in this paper. First the proposed algorithm provides a mathematical analysis method for the adaptive estimation of the high frequency contents based on the nature of the input audio. This method can be worked globally to any kind of audio such as speech and music that are restricted by sampling rate and bandwidth. We note that the method in paper [6] is based on the strict exponential modeling of DWT coefficients which is not appropriate to some of audio

cases and it even can cause some aliasing during the estimation procedure. Secondly, the proposed method can provide very robustic way of the estimation. For arbitrary level of DWT structure, we require only first level of DWT low frequency coefficients which assumed to be low sample rate audio for the estimation of the high frequency contents. Finally we would like to point out that the proposed algorithm can also work with any compressed audio that limits the bandwidth on the recovery, for example, 16Khz bandlimited audio of MP3.

3. EXPERIMENTAL RESULTS

In this section, we present some experimental results to demonstrate the performance of the proposed method. As a comparison purpose, a Corey's method in paper [6] is used as a reference. For the experiment, four different types of audio such as speech, electronic guitar drum and piano sound are considered. The recorded female speech is "We have just dock and a beautiful shores" at 16Khz sample rate, mono, 2.048 sec and downsampled to 8Khz in order to eliminate high frequency information from the source signal. On the other hand, the recorded electronic guitar, drum, and piano sound is sampled at 44.1Khz, stereo, 2.972 sec long, and downsampled to 22.05Khz for the high frequency elimination. These low sample rate audios are then spectrally enhanced to restore the sampling rate of 16Khz and 44.1Khz for speech and other sounds respectively by using the proposed method and a Corey's method. For the DWT analysis and synthesis, the Daubechie wavelet filter with filter length 30 is used for LPF and HPF.

Fig. 3 compares the experimental results for the female speech. Fig. 3(a) shows the spectrogram of the original speech at 16 Khz. The spectrograms of spectrally enhanced audio by the Corey's method and the proposed algorithm are shown in Fig. 3(b) and (c). Fig. 4 compares the experimental results for the electronic guitar sound From Fig. 3, we see some chopping and unstable high frequency recovery over 5 Khz in the spectrogram resulting from Corey's method while the proposed method shows quite good recovery of the high frequency contents of the original female speech. Similar observations can be made for the electronic guitar sound from Fig. 4. From these experimental results, we can confirm the stable and rich high frequency contents recovery for the proposed method.

4. CONCLUSION

In this paper, we present a new method for the high frequency enhancement of low sample rate audio. The algorithm is based on the adaptive LMS estimation of missing high frequency contents using the DWT expansion of low sample rate audio and the reconstruction by DWT synthesis. The proposed method can be worked globally to any kind of audio such as speech and music, especially for the audios that requires rich high frequency contents such as female speech, electronic guitar and drum sound. From the experimental results of spectrogram and sonic test, we confirm that the proposed algorithm works reasonably well and outperforms other algorithm.

References

[1] C. Avendaro, et. Al., "Beyond Nyquist: towards to recovery of broad-bandwidth speech from narrow-bandwidth speech", in Proc. Eurospeech, (Madrid), pp.165-168, 1995

[2] Y. M. Cheong, et., al, "Statistical recovery of wideband speech from narrowband speech", IEEE Trans. Speech Audio Process., vol.2, pp.544-548, 1994

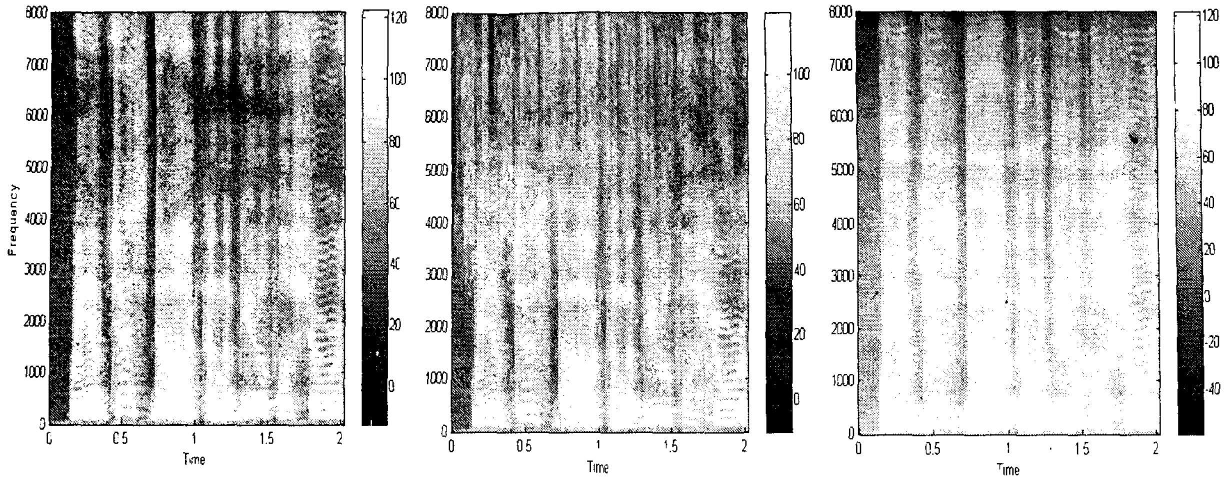
[3] J. Epps et. al., "A new technique for wideband enhancement of coded narrowband speech", Proc. IEEE Speech Coding Workshop (Porvoo Finland), 1999, pp. 174-176

[4] J. Valin, et. al., "Bandwidth extension of narrowband speech for low bit-rate wide band coding", IEEE Speech Coding Workshop, 17-20,Sep. 2000, pp. 130-132

[5] Chi-Min Liu, et. al., "High frequency reconstruction for band-limited audio signals", Proc. of 6th Conf. on Digital Audio Effects, London, UK, Sep. 8-11, 2003

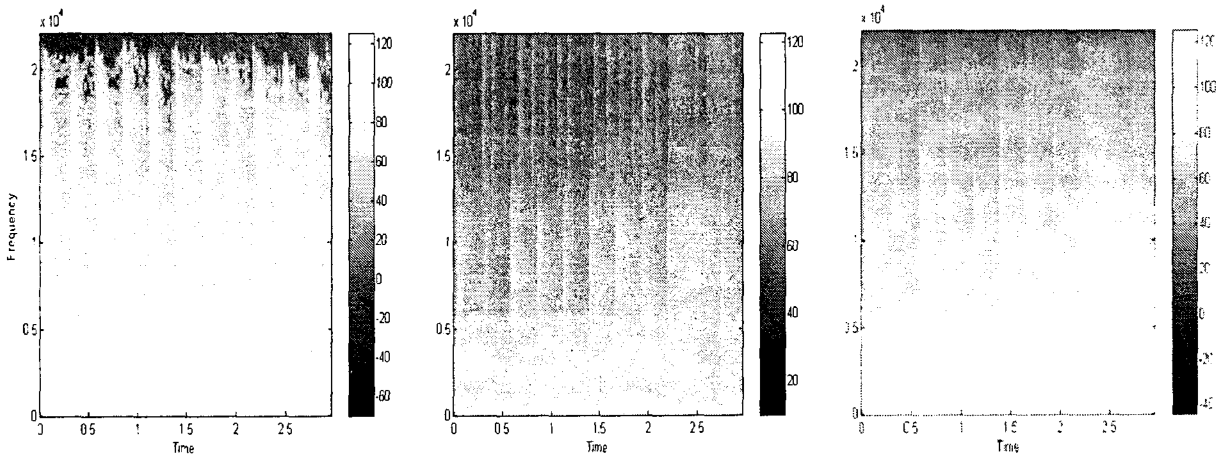
[6] Corey, C., "Wavelet signal processing of digital audio with applications in electro-acoustic music", Dartmouth College Master's Thesis, Dartmouth College, Hanover, New Hampshire:1996

[7] Strang, G. and Nguyen, T., Wavelets and filter banks, Wellesley-Cambridge, Wellesley, MA, 1996



(a) Original signal (b) Corey's method in [4] (c) Proposed method

Fig. 3 Spectrograms for: (a) original female speech at 16Khz sample rate, (b) high frequency enhanced with Corey's method in paper [6], (c) high frequency enhanced with the proposed method



(a) Original signal (b) Corey's method in [4] (c) Proposed method

Fig. 4 Spectrograms for: (a) original electronic guitar sound at 44.1Khz sample rate, (b) high frequency enhanced with Corey's method in paper [6], (c) high frequency enhanced with the proposed method