Audio Engineering Society

Convention Paper

Presented at the 123rd Convention
2007 October 5–8 New York, NY, USA

A Novel Audio Post-Processing Toolkit for the Enhancement of Audio Signals Coded at Low Bit Rates

Raghuram Annadana[1], Harinarayanan E.V[1], Deepen Sinha[1], and Anibal Ferreira[1, 2]

[1] ATC Labs, New Jersey, USA

[2] University of Porto, Portugal

Correspondence should be addressed to raghu@atc-labs.com

ABSTRACT

Low bit rate audio coding often results in the loss of a number of key audio attributes such as audio bandwidth and stereo separation. Additionally, there is also typically a loss in the level of details and intelligibility and/or warmth in the signal. Due to the proliferation, e.g. on Internet, of low bit rate audio coded using a variety of coding schemes and bit rates over which the listener has no control, it is becoming increasingly attractive to incorporate processing tools in the player which can ensure a consistent listener experience. We describe a novel post-processing toolkit which incorporates tools for (i) Stereo Enhancement, (ii) Blind Bandwidth Extension, (iii) Automatic Noise Removal and Audio Enhancement, and, (iv) Blind 2-to-5 channel upmixing. Algorithmic details, listening results, and audio demonstrations are presented.

1. INTRODUCTION

Audio coded at low bit rates exhibits loss with respect to a number of key audio attributes such as audio bandwidth, stereo separation and details in the signal. Another important issue with low bit rate coding is the degradation of audio quality due to noise. Sources of noise are plentiful varying from a live out-of-studio broadcast to scratchy vinyl records. Apart from disconcerting the listeners, a noisy audio, results in wastage of bit consumption at the encoder by naively coding noisy frames. Since an average consumer of digital audio (e.g., on Internet) comes across material coded at a variety of different bit rates and coding schemes there is desire to use player enhancements to ensure consistent audio listening experience.

We present details of a novel Audio Post-Processing Toolkit which consists of four audio playback enhancements tools

- Stereo Enhancement
- Blind Bandwidth Extension
- Automatic Noise Removal
- Blind 2-to-5 Channel Upmixing

Stereo signals with limited stereo are often found in the realm of compressed audio (i.e. music in compressed low bit rate formats) or in a FM receiver during instance of poor coverage (e.g. when the *Side* channel is lost or is of poor quality). The Stereo Enhancement tool enhances either a monophonic signal or a stereo signal with limited stereo image. Stereo signals with limited stereo are often found in the realm of compressed audio (i.e. music in compressed low bit rate formats). Our approach to stereo enhancement emphasizes image stability and creation of spaciousness across the entire frequency spectrum. Efforts are made to create stereo separation without adding any phasiness in dominant vocals. For signals with limited stereo optimum use of available stereo cues is made in generating the expanded stereo image. The algorithm is fully adaptive and is based on a detailed time-frequency analysis of the signals. It applies a combination of level and phase cues to achieve stereo separation and high image stability. The main strengths of the algorithm are accurate and detailed signal analysis and a novel scheme for the application of the phase cues.

Blind bandwidth extension is attractive in a number of situations. For example, a lot of MP3 material available on the Internet is coded at relatively low bit rates, e.g., 32-48 kbps whereby the codec preserves only about 8 kHz of audio bandwidth. Restoration of higher frequency components prior to playback can significantly enhance the listening experience. The advantage of a good blind Bandwidth Extension scheme is that it can be applied to degraded material over which the player has no control (as opposed to forward bandwidth extension schemes which require an analysis of original full bandwidth signal). A number of blind bandwidth extension schemes have been proposed [13, 14] but these are typically less than satisfactory because either the added high frequency components sound distorted or lack structure (e.g. are mostly noise like). We have developed a novel approach to blind bandwidth extension that sounds consistently good across a wide range of audio material and exhibits a significant amount of detail in the regenerated high frequency component. The scheme is an extension of our previously presented Fractal Self Similarity Model (FSSM) [1, 6] bandwidth extension technique, incorporated in the ATC Labs Audio Bandwidth Extension Toolkit (ABET) [1]. For the purpose of blind bandwidth extension, FSSM is applied to a modified signal domain with little or no spectral tilt. A different set of techniques are used to heuristically extend the spectral envelope. The resulting synthesized signal is a wide band signal with fairly detailed high frequency information.

Most of the commercial noise reduction techniques prudently address issues of quality degradation due to noise by effectively suppressing noise but at the expense of distorting the main signal. Also, another drawback with some of the noise reduction techniques is the requirement to select an appropriate noise profile or to run a two pass algorithm for selecting a suitable noise profile prior to the filtering stage. Such algorithms are less suitable for real time systems with a requirement of automatic wideband noise removal where the quality of audio is not to be compromised. The proposed algorithm addresses both these issues. Firstly, the technique achieves substantial noise reduction while preserving all the key audio characteristics of the primary signal sound and introduces minimal distortion to the primary sound. Secondly, identification of noise statistics is fully automatic and adaptive to any type of audio material.

In addition to the above mentioned tools, we have included our recently introduced method for blind upmix [10] in the toolkit as this can be used to vastly enhance a listener's experience.

The organization of the rest of the paper is as follows. Section 2 provides the algorithmic details of the stereo enhancement tool. Details of the Blind Bandwidth extension algorithm is provided in section 3. Section 4 and 5 provide details of the Noise Reduction and Blind Upmixing tools respectively. Some results are presented in section 6 and is followed by conclusions in section 7.

2. STEREO ENHANCEMENT

The input stereo signal is first partitioned into frames and each frame is successively analyzed to extract its temporal information using a Utility Filter Bank (UFB) [1]. The Utility Filter Bank (UFB) is an over-sampled complex modulated filterbank. A time frequency grid is available for further processing.

An overlap-add analysis is also performed on the input signal to calculate the ODFT [2] spectra of the audio. A downmix [10] is used in the ODFT domain to obtain a downmixed version of the input stereo. An accurate harmonic analysis [3, 10] is performed on the downmixed ODFT spectrum to reveal the harmonic structure in the signal. All dominant harmonic and all

its partials are marked on the bin scale. These marked bins are now mapped to the UFB band scale so as to obtain a new tonal map. The tonal map indicates UFB bands with vocal components. Care is taken to flag consecutive bands when partials are found to be very close to the band boundaries. A vocal preservation limit of 3500 Hz has been chosen empirically, i.e only partials below this limit are considered for the mapping.

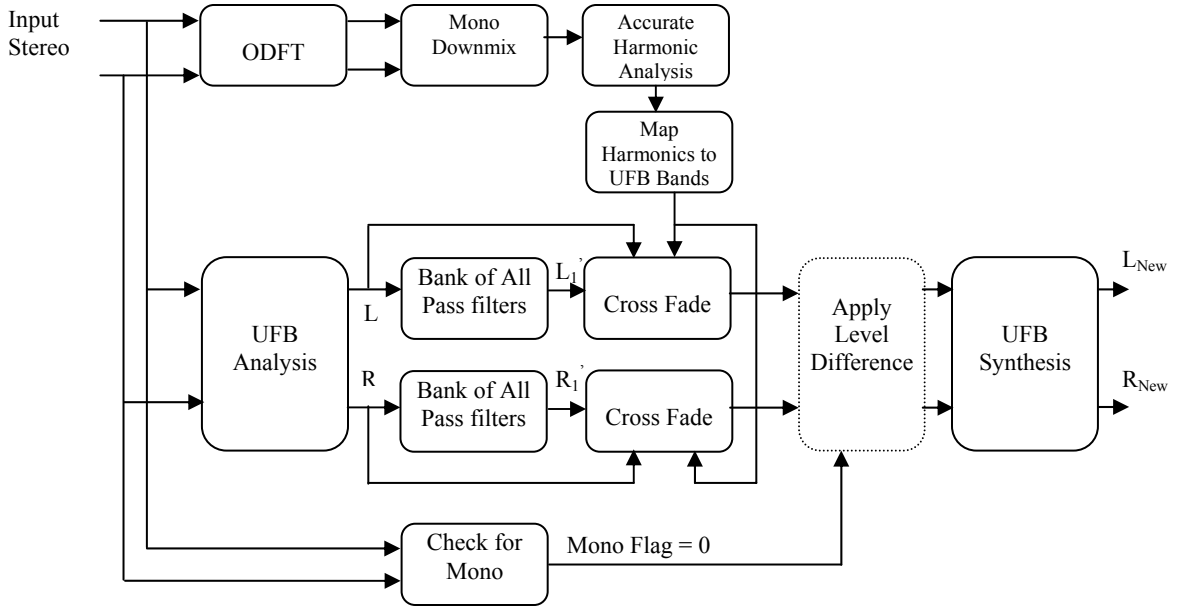

Figure 1: Architecture of the Stereo Enhancement Tool

Each time slice of the grid is now passed through a parallel bank of four all pass filters with varying delays and gains. The delays and gains have been empirically selected based on subjective listening tests. The signals from the parallel all pass filter bank are mixed so as to obtain two decorrelated versions of the audio L' and R'. One possible mixing matrix is shown below.

$$M = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \quad (1)$$

In order to avoid *phasiness* in the regenerated audio, only the time slices of the original stereo which have not been marked by the harmonic analysis module are now cross faded with the decorrelated signals L_1' and R_1'. A check is also made on the original subbands to test the nature of the original stereo. If the original audio was found to be a true mono signal (i.e. Left channel is identical to Right channel), then no further processing is performed for stereo regeneration.

If the signal was found to be stereo (but with stereo loss), then the level differences between the Left and Right channels are calculated. The level differences between the bands for frequencies greater than 4 kHz are modified if required based on the prior obtained statistics. An UFB synthesis is used to obtain the time domain version of the enhanced stereo.

3. BLIND BANDWIDTH EXTENSION

The blind high frequency reconstruction system accepts any low bit rate bandwidth constrained decoded audio as input and reconstructs the missing higher frequencies based on a detailed signal analysis.

Figure 2 shows the overall scheme for blind bandwidth extension. A Linear Predictive (LP) analysis is conducted on the input audio frame so as to obtain the narrowband LP coefficients utilizing Burg's method [4].

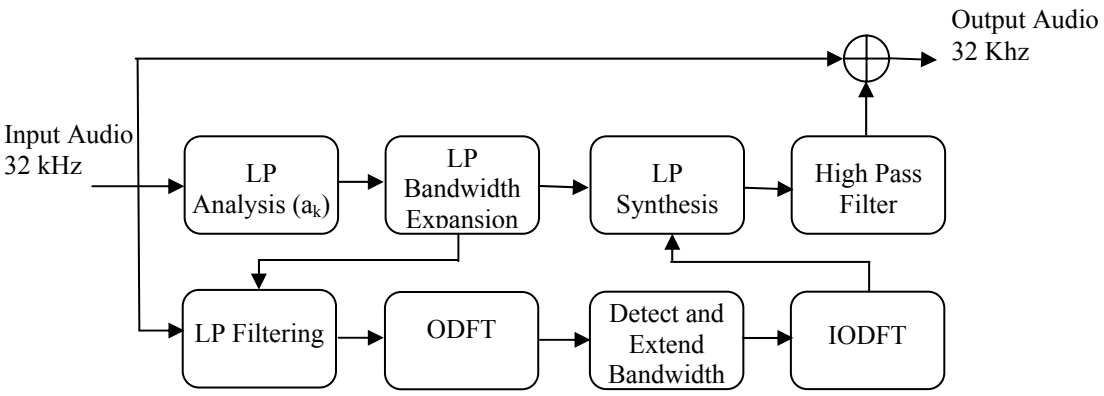

Figure 2: Architecture of Blind Bandwidth Extension

The scheme for bandwidth extension relies on (i) bandwidth expansion of flattened LP residuals using ABET tools, and, (ii) bandwidth extension of spectral envelope using through modification of LPC coefficients[5], Traditionally, bandwidth expansion of LPC coefficients has been used to improve stability of the LP filter [7], in this operation LP coefficients are modified according to Equation 2 below.

$$a_{new_i} = \gamma^i a_i; \qquad i = 1, 2, ..., M \tag{2}$$

where $\gamma < 1$ is a positive constant and M is the order of the filter.

The operation moves all the poles of the synthesis filter radially toward the origin, leading to improved stability. By doing so, the original spectrum is bandwidth expanded, in the sense that the spectrum becomes flatter, especially around the peaks, where the width is widened. Typical values for γ are between 0.988 and 0.996. However, for our application, we have used a value of γ which is around 0.6. The effect of this aggressively placed constant leads to a sufficiently softened audio after LP synthesis.

The bandwidth expanded coefficients are now used to filter the input audio so as to obtain an excitation. The ODFT of the LP residual is calculated. An automatic bandwidth detection procedure utilizing the energy of the bins as a measure is utilized to identify the maximum frequency content of the signal.

As noted above the bandwidth of LP residual is expanded using ABET tools. Specifically dilation and translation procedure inherent in the FSSM bandwidth extension scheme [1, 6] is used on the residual ODFT spectrum starting from the detected frequency cut off band to extend the bandwidth. It may noted that the FSSM parameters are heuristically computed (from the narrow band baseband) by emphasizing pitch continuity in the generation of the higher frequencies. Furthermore, since FSSM in this case is operating on a "flattened" signal no envelope adjustment is necessary as long as sufficient time resolution in the ODFT analysis is maintained (as noted in [6] envelope coherence between low and high frequencies is ensured by the FSSM operator). The bandwidth extended residual spectrum is transformed back to time domain using overlap-add IODFT synthesis. This in turn is fed into LP synthesis employing the bandwidth expanded LPCs as described above to obtain the output audio. A high pass filter is used to filter the signal above the detected frequency cut off. The original signal is added to the signal obtained from the high pass filter to obtain the bandwidth extended audio.

4. NOISE REDUCTION

The Automatic Noise Removal (ANR) algorithm attempts to remove wide-band background noise with the help of adaptive filtering techniques. The key distinguishing aspects of ANR are twofold. Firstly, it attempts to preserve the primary signal sound by performing a detailed harmonic analysis of the signal and by utilizing perceptual modeling and accurate signal analysis and synthesis, removes the primary signal components from the signal prior to the step of noise-removal. This helps preventing damage to the main signal sound due to the cross affect of the noise removal

algorithm (a key problem with most commercially available noise removal products). Secondly, ANR continuously and automatically updates noise profile/statistics with the help of a novel signal activity detection algorithm making the noise removal process fully automatic without any human intervention. The basic block diagram of the noise removal technique is as shown in Fig.1. The core of the noise removal algorithm is built around a de-noising Kalman filter. Further details of filter design, filter parameter estimation and principal signal preservation are discussed further at [8].

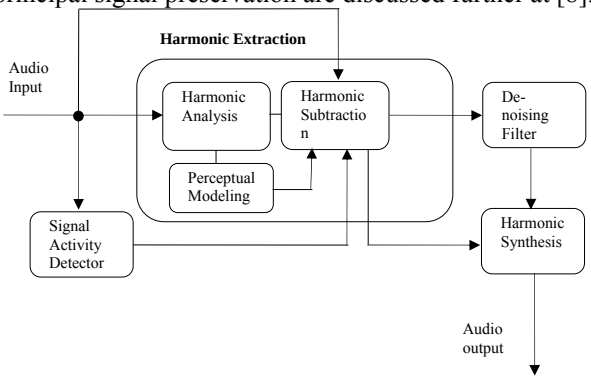

Figure 3: Architecture of ANR

5. BLIND UPMIX

A new blind upmixing approach was developed as a part of the Immersive Soundfield Rendition (ISR) parametric multi-channel coding platform [9, 10]. This tool has also been incorporated in the post processing toolkit. The entire scheme is briefly described in this section.

The front and center channels are generated by performing a Principal Component Analysis (PCA) of the input stereo which produces two vectors indicating the direction of both a dominant signal and the remaining signal. The surround channels are synthesized by adaptively mixing three components viz. an ambience signal extracted from the stereo audio, a second reverberation component derived from the dominant principal component vector and a third Image Movement Spatialization Component (*IMSC*) based on the original carrier audio. The third, *IMSC*, component which is a weighted version of the original carrier audio is based on the rate of change of the stereo angle of the dominant signal component. Synthesizing the *IMSC* component involves exploiting the stereo image angle (α) of the dominant signal component that is available as a byproduct of the Principal component analysis used for the generation of front channels. The rate of change of the stereo image angle gives a measure of the movement of the stereo image over the front plane in the stereo downmix. This information is utilized in the final mixing process for the creation of the synthetic surround allowing for a tangible shift in the surround audio image from front to back (or back to front), thereby increasing the perception of spaciousness in the synthetic multi-channel signal. This space filling effect creates a pleasant perception. The final surround audio is synthesized by mixing Ambience (complex audio content), reverberation of the dominant component (naturalness) and the IMSC component.

6. RESULTS

To characterize the performance of the various tools described, various informal subjective and objective tests were conducted. More formal subjective tests are in progress. A group of five expert listeners was used for subjective quality assessment of the tools. In the case of the stereo enhancement and blind bandwidth extension tools, MP3 coded audio was used for verification. All audio material was sourced from commercially available music CDs and was encoded at bit rates of 56 kbps onwards. About 30 minutes of audio were used for testing. All audio samples were coded with intensity joint stereo and mid side joint stereo settings of the MP3 encoder. All audio samples were played back on Sennheiser HD650 headphones in addition to playing the same samples in a sound proofed multichannel listening environment.

The original MP3 coded sample was first played and only after the listener was reasonably confident of its stereo qualities the enhanced version was played back. Additionally, the stereo expander tool in Adode Audition 1.5 was used as an anchor. Listeners pointed out the *phasiness* as a main issue with samples from Adobe Audition, in addition to a loss of the vocal component in music. All listeners labeled a preference to our enhanced version. Some of the attributes labeled for the enhanced version included spacious, clear, filling, warm, distinct and wide. Listeners also noted a preservation of the dominant vocal component in music and an absence of *phasiness*.

In the case of blind bandwidth extension tool, the listeners readily labeled a preference for bandwidth extended audio coded at 56 and 64 kbps MP3 coding (where the MP3 encoder was forced to encode only 8kHz of audio bandwidth and the blind bandwidth

extension tool was used to extend the bandwidth to 15 kHz). In another test for this tool, clean PCM audio (uncompressed) was band limited to 6 - 8 kHz and the blind bandwidth extension tool was applied to extend the bandwidth to 15 kHz. In every case the listeners exhibited a clear preference for the extended audio. In order to further evaluate the subjective quality, objective scores were obtained from the PEAQ [12] audio quality evaluation algorithm. Audio coded at 56 kbps with bandwidth extended from 8 kHz onwards was used. Individual PEAQ scores were obtained by testing the original audio compared against the MP3 coded audio and original audio compared with bandwidth extended audio. An average increase of 0.31 was seen on the PEAQ measurement scale.

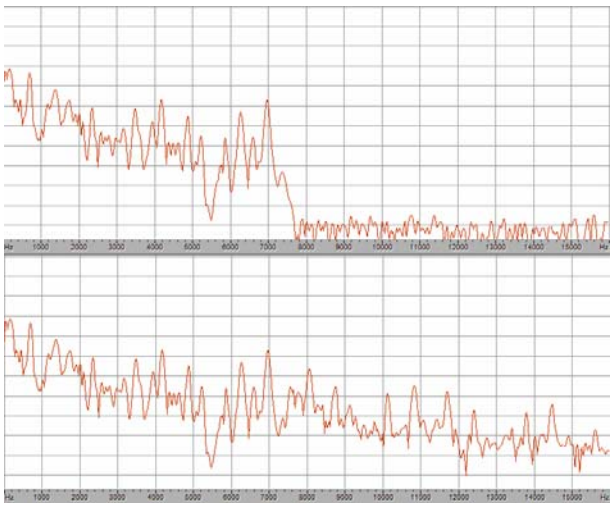

Figure 4: Original Audio and Bandwidth Extended Audio from 7.5k Hz onwards

In the evaluation of the noise reduction tool, a comparison was made against some of the commercially available noise removal products like Dart XP Pro V1.1.6p, Adobe Audition 1.5 and GoldWave v5.19. The database used for this test were from various real life broadcast and communication setups with a high level of background noise with varying characteristics. Listeners were asked to provide a score on the scale of 1 to 5 with 5 being the highest quality noise removal for various noise removal algorithms. The scores were based on factors like effectiveness of noise removal, distortion introduced after processing and the clarity/naturalness of the audio. The graph plotted in Fig. 4 shows the ratings of this experiment. ANR performed well in suppressing noise and also introduced the least distortion in the processed audio by preserving main signal components. A slightly elaborate analysis of this result can be found at [8].

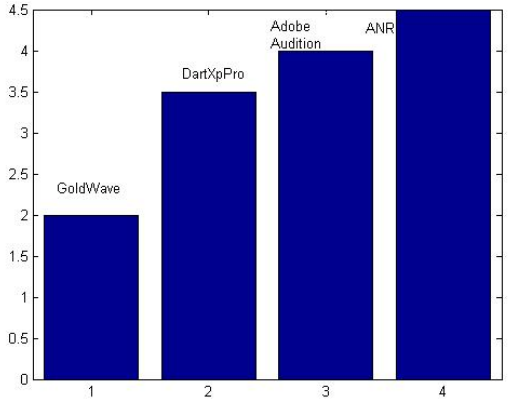

Figure 4: Subjective analysis of commercial noise removal modules

Listeners reported an increased image quality and accuracy in the case of the blind upmix algorithm [10].

7. CONCLUSIONS

We have presented the architectures and the operation of the various tools in our post processing toolkit. Four different tools have been discussed viz. stereo enhancement, blind bandwidth extension, noise reduction and blind upmix. The stereo enhancement tool can be used in conjunction with the blind bandwidth extension tool to effectively regenerate stereo from monophonic signals and reconstruct high missing frequencies. In addition, we have presented techniques to prevent *phasiness* in the dominant vocals while enhancing stereo. Techniques have also been presented to preserve the primary signal audio in noise removal utilizing detailed signal analysis. The recently introduced blind upmix scheme which adaptively mixes three components – viz. ambience, reverberated dominant PCA signal and the original carrier has also been incorporated in the toolkit. Objective and subjective tests have also been presented and they indicate an improvement in the perceived audio.

8. REFERENCES

[1] D. Sinha, A. J. S Ferreira and Harinarayanan E. V., "*A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)*", in the preprints of 120th

Convention of the Audio Engineering Society, May 2006.

[2] Anibal J. S. Ferreira, "*Spectral Coding and Post-Processing of High Quality Audio"*, Ph.D thesis, Faculdade de Engenharia da Universidade do Porto-Portugal, 1998.

[3] Anibal J. S. Ferreira, "*Accurate Estimation in the ODFT Domain of the Frequency, Phase and Magnitude of Stationary Sinusoids*", in 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October 21-24 2001, pp. 47-50

[4] J. P. Burg, "*Maximum Entropy Spectrum Analysis*", Ph. D. Dissertation, Stanford Univ., 1975.

[5] Peter Kabal, "*Ill-Conditioning and Bandwidth Expansion in Linear Prediction of Speech*", Proc IEEE Int. Conf. Acoustics, Speech, Signal Processing (Hong Kong), pp. I-824-I-827, April 2003.

[6] D. Sinha, A. J. S Ferreira and D. Sen, "*A Fractal Self-Similarity Model for the Spectral Representation of Audio Signals*", in the preprints of 118th AES Convention, Barcelona, Spain.

[7] Wai C. Chu, "*Speech Coding Algorithms*", John Wiley & Sons, 2003.

[8] Harinarayanan E.V, Deepen Sinha, Shamail Saeed, and Anibal Ferreira, "*A Novel Automatic Noise Removal Technique for Audio and Speech Signals*", to be presented in the 123rd AES Convention, New York, October, 2007.

[9] Chandresh Dubey, Richa Gupta, Deepen Sinha and Anibal Ferreira, "*A Novel Very Low Bit Rate Multi-Channel Audio Coding Scheme Using Accurate Temporal Envelope Coding and Signal Synthesis Tools*", in the preprints of 121st AES Convention, October 5-8, 2006 San Francisco, CA, USA.

[10] Chandresh Dubey, Raghuram A., Deepen Sinha and Anibal Ferreira, "*New Enhancements to Immersive Sound Field Rendition (ISR) System*", in the preprints of 122nd AES Convention, Vienna, Austria, May 2007.

[11] Anibal J. S. Ferreira, "*Perceptual Coding of Harmonic Signals*", in the preprints of 100th Convention of the Audio Engineering Society, May 2005, 1996.

[12] Thilo Thiede, William C. Treurniet, Roland Bitto, Christian Schmidmer, Thomas Sporer, John G. Beerends, Catherine Colomes, "*PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality*", JAES Vol. 48 No. 1/2 pp. 3-29; Jan/Feb 2000

[13] Erik Larsen, Ronald Aarts and Micheal Danessis, "*Efficient high-frequency bandwidth extension of music and speech*", in the preprints of 112th Convention of the Audio Engineering Society, May 2002, Munich, Germany, 1996.

[14] Manish Arora, Joonhyun Lee, and Sangil Park, "*High Quality Blind Bandwidth Extension of Audio for Portable Player Applications*", in the preprints of 120th Convention of the Audio Engineering Society, May 20-23, 2006, Paris, France,.