Decimation-Whitening Filter in Spectral Band Replication

Han-Wen Hsu and Chi-Min Liu

***Abstract*—MPEG-4 High-Efficiency Advanced Audio Coding (HE-AAC) has adopted spectral band replication (SBR) to efficiently compress the high-frequency part of the audio. In SBR, linear prediction is applied to low-frequency subbands to suppress tonal components and smooth the associated spectra for replicating to high-frequency bands. Such a tone-suppressing process is referred to as whitening filtering. In SBR, to avoid the alias artifact incurred by spectral adjustment, a complex filterbank instead of real filterbank is adopted. For QMF subbands, this paper demonstrates that the linear prediction defined in the SBR standard results in a predictive bias. A new whitening filter, called the decimation-whitening filter, is proposed to eliminate the predictive bias and provide advantages in terms of noise-to-signal ratio measure, frequency resolution, energy leakage, and computational complexity for SBR.**

***Index Terms*—analytic signal, bandwidth extension, high-frequency (HF) reconstruction, linear prediction, spectral band replication (SBR), whitening.**

I. Introduction

SPECTRAL band replication (SBR) [1]–[3] has been introduced into MPEG-4 High-Efficiency Advanced Audio Coding (HE-AAC) [2] as a bandwidth-extension technique. In the QMF subband domain, SBR reconstructs high-frequency (HF) signals by transposing and adjusting replicated low-frequency (LF) signals thanks to the strong correlation of spectral harmonic characteristics. Since the SBR technique requires significantly lower bit rate to code the high bands and reduces the AAC coder bandwidth, the AAC encoder can compress the LF part with most of the available bits to achieve high coding gain.

Rather than the cosine modulated filterbank (CMFB) commonly employed in audio coding, SBR utilizes the comparatively high-complexity complex-exponential modulated filterbank (CEMFB) [1] to eliminate the main alias terms and thus avoid the alias artifact introduced from spectral adjustment or equalization. On the other hand, the tonal components existing in the replicated LF bands may bring undesired artifacts like tonal spikes [4] into the reconstructed HF bands. Therefore, inverse filtering based on the second-order linear prediction (LP) [2] is provided to whiten the replicated LF bands. This paper, however, demonstrates that the conventional LP method defined in the SBR standard has natively a predictive bias which affects the whitening effect and the noise-to-signal ratio (NSR) measure.

Manuscript received October 24, 2009; revised July 03, 2010, November 27, 2010; accepted March 08, 2011. Date of publication March 22, 2011; date of current version August 19, 2011. This work was supported by the National Science Council under Contract NSC99-2221-E-009-113-MY2. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Susanto Rahardja.

The authors are with the Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan. (e-mail: hwhsu@csie.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2011.2131130

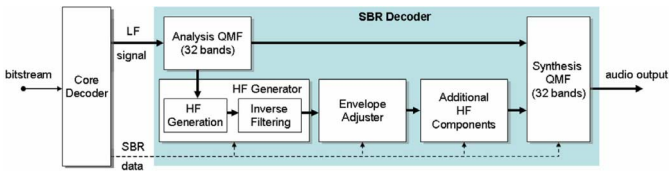

Fig. 1. Block diagram of the SBR decoder.

This paper is organized as follows. Section II provides an overview of SBR and models the CEMFB subbands as analytic signals for later analysis. In Section III, the predictive bias is demonstrated through the first-order and second-order autoregressive (AR) modeling on analytic signals together with the empirical verification on the CEMFB subbands. Section IV proposes an alternative whitening filter for removing the bias for the SBR algorithm. Section V concludes this paper.

II. Preliminaries

A. Overview of SBR

The SBR is a technique of bandwidth extension or high-frequency reconstruction and can be combined with any audio core coder such as AAC and MP3 (MPEG-I Layer-III). Only a small amount of side information, including spectral envelope data and control parameters for additional means such as inverse filtering and noise/sinusoidal addition, is transmitted from the encoder to the decoder to guide the high-frequency reconstruction.

As depicted in Fig. 1, the SBR decoding has three major procedures. In the HF generator, the low bands split from a decoded LF signal are first transposed to HF. Subsequently, the inverse filtering is applied to the regenerated HF bands to suppress the undesired sinusoidal components from the replicated LF bands and thus control tonality. The inverse filtering is performed by in-band filtering using an adaptive spectral whitening filter. The second-order covariance method is employed to evaluate the whitening filters on the low bands. Furthermore, to control the amount of inverse filtering, the chirp factor given from the bitstream is used to move the zeros of the LP filters toward the origin. The regenerated high band $x_k[n]$ for QMF subband k and time slot n is defined as

$$x_k[n] = x_l[n] - a_l(1) \cdot c_k \cdot x_l[n-1] - a_l(2) \cdot c_k^2 \cdot x_l[n-2]$$

where $a_l(1)$ and $a_l(2)$ are the predictive coefficients estimated on the low band $x_l[n]$, and c_k is the chirp factor whose range

is between 0 and 0.98. In the envelope adjuster, the envelope of the regenerated high bands is scaled according to the transmitted spectral envelope information that is represented by the average energies in time–frequency grids. Subsequently, additional tones and random noise are compensated to adjust the tonality of the reconstructed high bands. Finally, all the low and high bands are synthesized to generate a full-bandwidth decoded signal. More details about the SBR algorithm can be found in [1]–[3].

B. CEMFB Subbands and Analytic Signals

The discrete-time analytic signal $x_+[n]$ corresponding to a real signal $x[n]$ [5] is defined as $x[n]+j\hat{x}[n]$, where $\hat{x}[n]$ denotes the discrete-time Hilbert transform of $x[n]$

$$\hat{x}[n] = \sum_{k=-\infty, k\neq 0}^{\infty} \frac{2}{\pi} \cdot \frac{\sin^2(\pi k/2)}{k} \cdot x[n-k]. \tag{1}$$

In the frequency domain, the relation between the original and analytic signals is given by

$$X_+(\omega) = \begin{cases} 2X(\omega), & \forall 0 < \omega < \pi \\ 0, & \forall -\pi < \omega < 0 \end{cases}. \tag{2}$$

Similarly, the analytic signal $x_-[n]$ containing merely the negative spectrum can be defined as $x[n] - j\hat{x}[n]$. For convenience, we call $x_+[n]$ and $x_-[n]$ the "positive" and "negative" analytic signals, respectively.

Both the analysis and synthesis filters of the 64-channel CEMFB system used in SBR are defined by

$$h_k[n] = f_k[n] = p[n] \cdot \exp\left(\frac{j\pi}{2M}(2k+1)(n-N/2)\right) \tag{3}$$

for $k = 0, \ldots, M-1$, where M is the number of channels, and N is the order of the prototype filter $p[n]$. Compared with the CMFB, the CEMFB adds an imaginary part that consists of sine modulated versions of the same prototype filter, which can be interpreted as the Hilbert transform of the real part. Accordingly, the resultant subbands decimated by M can be approximately regarded as the analytic signals of the real output obtained from the CMFB [1]. Moreover, the CEMFB subbands alternately consist of positive and negative analytic signals.

In the absence of either the positive or negative side band, the excitation noise for a CEMFB subband can be also regarded as an analytic signal which has flat power spectrum density (PSD) in the other side band; but it is no longer white. Nevertheless, the whiteness of excitation noise is a desirable property for confirming the asymptotically unbiased LP estimation of spectral peaks [6]. This non-whiteness implies that the absence of one side band leads to a predictive bias which is demonstrated in Section III.

III. Linear Predictive Bias on Analytic Signals

This section demonstrates and quantifies the predictive bias of the first-order and second-order LPs on analytic signals. We first analyze the bias from the theoretical derivation on ideal analytic signals. Next, we confirm through empirical verification the bias on the CEMFB subbands which are generated by the modulated non-ideal prototype filter. The affection of the bias in SBR will be discussed in Section IV. The derivation and illustration in this section and Section IV are given according to the positive analytic signal model, and the same result can be extended to the negative one.

A. First-Order LP on Analytic Signals of First-Order AR Model

Consider the analytic signal modeled by the AR model with single pole $r_0 e^{j\theta_0}$ in the frequency domain

$$X_+(\omega) = \frac{E(\omega)}{1 - r_0 e^{j\theta_0} e^{-j\omega}} \tag{4}$$

where the PSD of the excitation signal $E(\omega)$ is assumed to be 1 for $0 < \omega < \pi$ and 0 for $\pi < \omega < 2\pi$, and the pole locates inside the upper half of the unit circle. The mean-square error function of the first-order predictive filter $re^{i\theta}z^{-1}$ on the single-pole analytic signal is expressed as

$$F(r,\theta) = \int_0^{\pi} \left| \frac{1 - re^{j\theta}e^{-j\omega}}{1 - r_0 e^{j\theta_0} e^{-j\omega}} \right|^2 d\omega \tag{5}$$

$$\text{or} \quad F(r,\theta) = \int_0^{\pi} \frac{1 - 2r\cos(\omega-\theta) + r^2}{1 - 2r_0\cos(\omega-\theta_0) + r_0^2} d\omega. \tag{6}$$

The minimum mean-square error (MMSE) predictive filter can be obtained by solving two equations: $\partial F/\partial\theta = 0$ and $\partial F/\partial r = 0$. Thus, in polar coordinates, the conditions of the zero position $(\tilde{r}, \tilde{\theta})$ of the MMSE filter can be derived as

$$A(\tilde{\theta}) = \int_0^{\pi} \frac{\sin(\omega - \tilde{\theta})}{1 - 2r_0\cos(\omega-\theta_0) + r_0^2} d\omega = 0 \tag{7}$$

$$B(\tilde{r}, \tilde{\theta}) = \int_0^{\pi} \frac{\cos(\omega - \tilde{\theta}) - \tilde{r}}{1 - 2r_0\cos(\omega-\theta_0) + r_0^2} d\omega = 0. \tag{8}$$

By using the trigonometric properties $\sin(\alpha - \beta) = \sin(\alpha)\cos(\beta) - \cos(\alpha)\sin(\beta)$ and $\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$, (7) and (8) can be rewritten as

$$A(\tilde{\theta}) = \cos(\tilde{\theta} - \theta_0)S - \sin(\tilde{\theta} - \theta_0)C = 0 \tag{9}$$

$$B(\tilde{r}, \tilde{\theta}) = \cos(\tilde{\theta} - \theta_0)C + \sin(\tilde{\theta} - \theta_0)S - \tilde{r}K = 0 \tag{10}$$

where three integrations S, C, and K are defined as

$$S = \int_0^{\pi} \frac{\sin(\omega - \theta_0)}{1 - 2r_0\cos(\omega-\theta_0) + r_0^2} d\omega \tag{11}$$

$$C = \int_0^{\pi} \frac{\cos(\omega - \theta_0)}{1 - 2r_0\cos(\omega-\theta_0) + r_0^2} d\omega, \tag{12}$$

$$K = \int_0^{\pi} \frac{1}{1 - 2r_0\cos(\omega-\theta_0) + r_0^2} d\omega. \tag{13}$$

Thus, from (9), the angle of the zero of the MMSE predictive filter is given as

$$\tilde{\theta} = \theta_0 + \arctan\left(\frac{S}{C}\right). \tag{14}$$

By substituting (14) into (10), the radius of the zero of the MMSE predictive filter is derived as

$$\begin{aligned}\tilde{r} &= \frac{\cos\left(\arctan\left(\frac{S}{C}\right)\right)\cdot C + \sin\left(\arctan\left(\frac{S}{C}\right)\right)\cdot S}{K} \\ &= \frac{\sqrt{C^2+S^2}}{K}.\end{aligned} \tag{15}$$

In the derivation of (14) and (15), we use the fact that C and $\tilde{r}$ are always positive to select the correct value of $\tilde{\theta}-\theta_0$. Also, when $r_0 > 0$, the closed forms of the three integrations are given as (see Appendix A)

$$S = \frac{1}{2r_0}\log\left(\frac{1+2r_0\cos(\theta_0)+r_0^2}{1-2r_0\cos(\theta_0)+r_0^2}\right) \tag{16}$$

$$C = \frac{\left(1+r_0^2\right)K-\pi}{2r_0} \tag{17}$$

$$\begin{aligned}K = \frac{1}{1-r_0^2}\Bigg[&T + \arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0)+r_0^{-1}}\right) \\ &-\arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0)-r_0^{-1}}\right)\Bigg],\end{aligned} \tag{18}$$

where we have the equation as shown in (19) at the bottom of the page.

1) General Case of $r_0 > 0$: Equation (14) shows an angle bias, $\arctan(S/C)$, between $\tilde{\theta}$ and θ_0, which is nonzero except the case of $\theta_0 = \pi/2$ Moreover, from the integration in (7), it can be shown that $A(\theta_0)\cdot A(\pi/2) < 0$ except $\theta_0 = \pi/2$. By Root Location Theorem, the root of (7) locates within the open interval between θ_0 and $\pi/2$. Therefore, in general, the angle of the zero of the MMSE filter is biased from θ_0 toward $\pi/2$ and cannot match the pole $r_0 e^{i\theta_0}$ of the AR model. Fig. 2 illustrates that the angle bias with $r_0 = 1/2$ increases as θ_0 is far away from $\pi/2$. By substituting $\theta_0 = \pi/2$ into (16)–(18) and using the trigonometric properties $\arctan(-\beta) = -\arctan(\beta)$ and $\arctan(1/\alpha) = \pi/2 - \arctan(\alpha)$ for $\alpha > 0$, it can be derived that

$$\tilde{r} = r_0 + \frac{1-r_0^2}{r_0}\cdot\frac{2\arctan(r_0)}{\pi+4\arctan(r_0)}. \tag{20}$$

Although $\tilde{\theta}$ is fitted to $\theta_0 = \pi/2$, there exists a bias between $\tilde{r}$ and r_0 in (20). Fig. 3 depicts the radius curve corresponding to

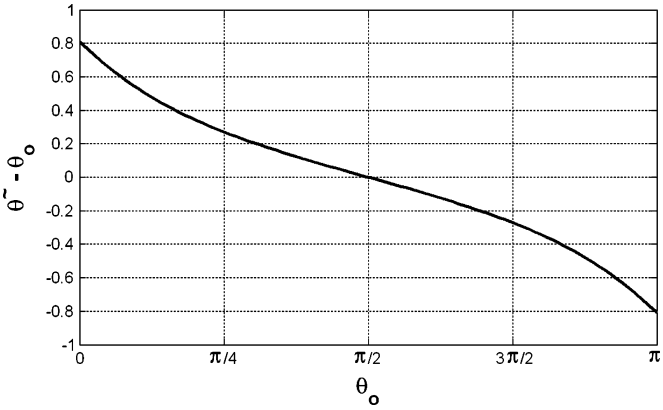

Fig. 2. Angle biases for different θ_0 values with $r_0 = 1/2$.

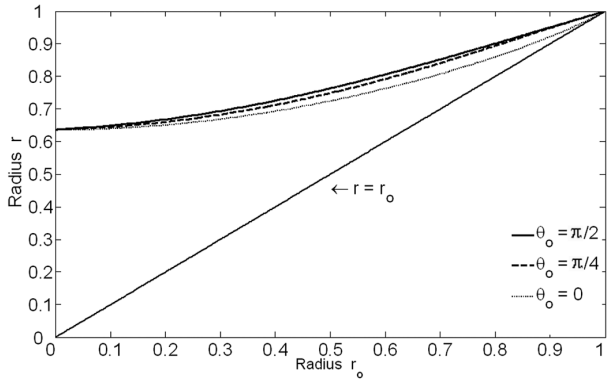

Fig. 3. Radius of the zero of the MMSE predictive filter on single-pole analytic signals with different θ_0.

$\theta_0 = \pi/2$ together with those corresponding to $\theta_0 = \pi/4$ and 0. The curves show that the radius bias increases as the pole of the predicted spectrum moves away from the unit circle. This trend implies that the prediction on noise-like signals should have a larger radius bias than that on tonal signals.

2) Flat-Spectral Case of $r_0 = 0$: Corresponding to a real white-spectral signal, e.g., an impulse signal or a white noise signal, the analytic signal can be modeled by (4) with $r_0 = 0$. Substituting $r_0 = 0$ into (7) yields $\cos(\pi - \tilde{\theta}) = \cos(\tilde{\theta})$; thus, $\tilde{\theta}$ should be $\pi/2$ or $3\pi/2$. Similarly, solving (8) with $r_0 = 0$ leads to $\tilde{r} = 2\sin(\tilde{\theta})/\pi$. Since $\tilde{r}$ is non-negative, the zero of the MMSE filter on the analytic signal positions at $(\tilde{r}, \tilde{\theta}) = (2/\pi, \pi/2)$, instead of the origin. Furthermore, the MMSE is $F(2/\pi, \pi/2) = \pi - 4/\pi$, and the estimated NSR is $(\pi - 4/\pi)/\pi \approx 0.594$ which is much lower than 1 that is the expected NSR value. Fig. 4 illustrates the first-order whitening processing on four analytic signals. In the absence of the negative bands, all

$$T = \begin{cases} \pi - \arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0)+r_0}\right) + \arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0)-r_0}\right), & |\cos(\theta_0)| > r_0 \\ \frac{3\pi}{2} + \arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0)-r_0}\right), & \cos(\theta_0) = -r_0 \\ 2\pi - \arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0)+r_0}\right) + \arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0)-r_0}\right), & |\cos(\theta_0)| < r_0 \\ \frac{\pi}{2} - \arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0)+r_0}\right), & \cos(\theta_0) = r_0 \end{cases} \tag{19}$$

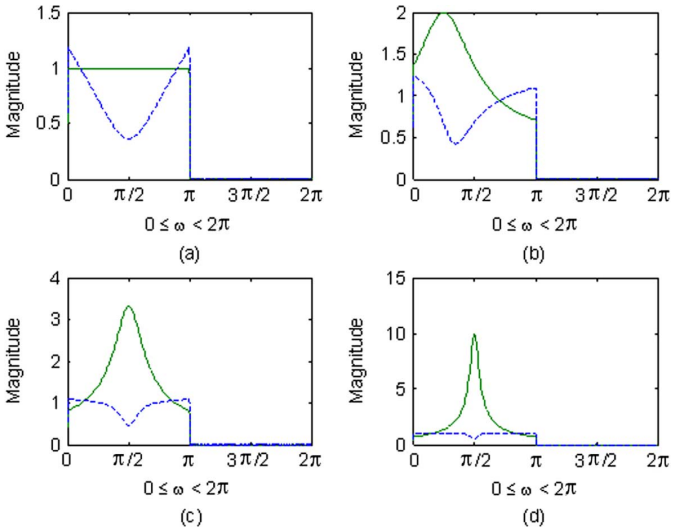

Fig. 4. Whitening processing on analytic signals of first-order AR model by first-order LP. (a) Flat-spectral analytic signal, (b)–(d) single-pole analytic signals with $(r,\theta) = (0.5, \pi/4), (0.7,\pi/2)$ and $(0.9,\pi/2)$. The zero location (r,θ) of first-order whitening filter in (a)–(d) are (0.6369, $\pi/2$), (0.7720, 0.3363π), (0.8594,$\pi/2$), and (0.9510, $\pi/2$), respectively. (Solid line: the original signals, dashed line: the whitened signals; these simulations are implemented via 2048-point DFT.) For ensuring the orthogonality of the real and imaginary parts of the analytic signals simulated by discrete Fourier transform (DFT) [6], the frequency response of the excitation signal at $\omega = 0$ and π is 1/2, in stead of 1.

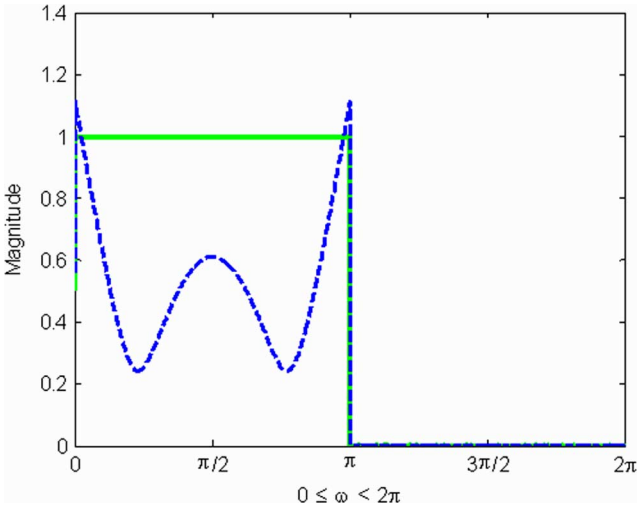

Fig. 5. Whitening processing on the flat-spectral analytic signal by second-order LP. The estimated NSR value is 0.3181, and the two zeros position at $(r,\theta) = (0.8257, 0.2247\pi)$ and (0.8257, 0.7753 π). (Solid line: original signals, dashed line: whitened signals; the simulation is implemented via 2048-point DFT and covariance method.)

the whitened analytic signals have additional spectral hollows in the positive bands.

B. Second-Order LP on Flat-Spectral Analytic Signals

The mean-square error function $F(r_1, r_2, \theta_1, \theta_2)$ of the second-order LP filter on the analytic signal corresponding to a white-spectral signal is expressed as

$$\int_0^{\pi} |(1 - r_1 e^{j\theta_1} e^{-j\omega})(1 - r_2 e^{j\theta_2} e^{-j\omega})|^2 d\omega \qquad (21)$$

where the PSD of the analytic signal is assumed to be 1 for $0 < \omega < \pi$ and 0 for $-\pi < \omega < 0$. As shown in Appendix B, the radii and angles of the two zeros of the MMSE filter are given by

$$\tilde{r}_1 = \tilde{r}_2 = \frac{2}{\sqrt{\pi^2 - 4}} \approx 0.826 \qquad (22)$$

$$\tilde{\theta}_1 = \arcsin\left(\frac{\pi}{2\sqrt{\pi^2 - 4}}\right) \approx 0.2245\pi \quad \text{and } \tilde{\theta}_2 \approx 0.7755\pi. \qquad (23)$$

The MMSE is $\pi(\pi^2-8)/(\pi^2-4) \approx 0.3185\pi$, and the estimated NSR is about 0.317, which is lower than the one evaluated by the first-order LP. Fig. 5 shows the resultant spectral hollows on the flat-spectral analytic signal by the second-order LP.

C. Empirical Verification for SBR

An empirical example is conducted in Fig. 6 for the first-order and second-order LPs by the covariance method. In the example, the original signal is a 32-point CEMFB subband signal of an impulse. As can be seen, the spectral hollows are shaped on the whitened signals in the frequency domain. For the first-order case in Fig. 6(a), the radius and angle of the zero of the LP filter are 0.5676 and $\pi/2$, and the estimated NSR value is 0.6778. For the zeros of the second-order LP filter in Fig. 6(b), their common radius is 0.6891 and their angles are 0.2078 π and 0.7922 π, respectively; the estimated NSR value is 0.5249. Accordingly, we can expect that the estimated NSR values in SBR for white-spectral or noise-like signals will be underestimated by about 30% and 50% with the first-order and second-order LPs, respectively. Through the above theoretic analysis on the ideal analytic signal model, we can also expect that the predictive bias becomes significant as the NSR of the predicted spectrum increases. This result is different from the intuition that the inverse filter should keep or slightly shape the spectra of noise-like signals.

IV. Decimation-Whitening Filter

As shown above, the non-whiteness of the excitation noise components in analytic signals results in the predictive bias. To remove the non-whiteness, the decimation by two should be applied to the CEMFB subbands before the covariance method is performed. The new approach has benefits in terms of frequency resolution, NSR measure, energy leakage, and computational complexity.

A. Decimation-Whitening Filter for SBR

The relation between the original analytic signal and the decimated signal by two is expressed in the frequency domain as

$$X_d(2\omega) = X(\omega) \qquad (24)$$

for either $0 < \omega < \pi$ or $-\pi < \omega < 0$, where X and X_d denote the Fourier transforms of the analytic and the decimated signals, respectively, and the range of ω depends on the absent side band of the analytic signal. Applying a second-order parameter

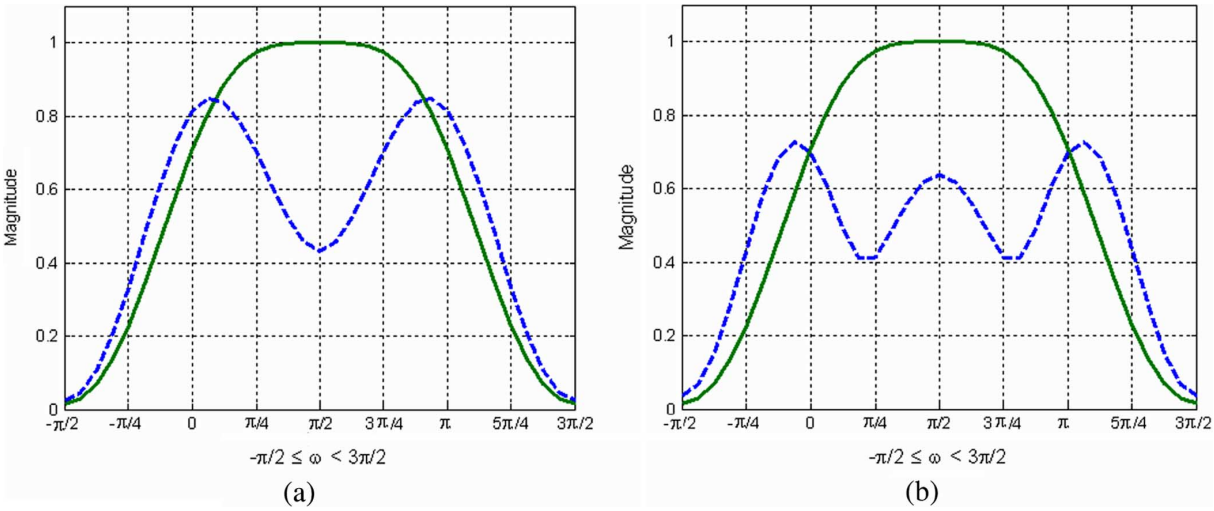

Fig. 6. Whitening processing on a 32-point CEMFB subband signal of an impulse. (a) First-order LP. (b) Second-order LP. (Solid line depicts original signals; dashed line depicts whitened signals.)

estimation method to the decimated signal can obtain two predictive coefficients a_1 and a_2, and then the estimated PSD of the analytic signal is given as

$$P_X(\omega) = \frac{\sigma_e^2}{|1 - a_1 e^{-j2\omega} - a_2 e^{-j4\omega}|^2} \tag{25}$$

for either $0 < \omega < \pi$ or $-\pi < \omega < 0$, where σ_e^2 denotes the variance of the residuals. Consequently, for the analytic signal, the fourth-order LP filter derived from the second-order LP filter of the decimated signal can be given in the z-transform as

$$H(z) = 1 - a_1 z^{-2} - a_2 z^{-4}. \tag{26}$$

Whitening filtering can be interpreted as putting zeros to the z-domain. In the SBR standard, the chirp factor is used to move the positions of the zeros toward the origin in the z-domain. The chirp factor can be combined with the fourth-order LP filter in the following way:

$$\hat{H}(z) = 1 - a_1 \cdot c^2 \cdot z^{-2} - a_2 \cdot c^4 \cdot z^{-4} \tag{27}$$

where c denotes the chirp factor. Although moving the zeros toward the origin reduces the whitening effect on a tonal signal, the chirp factor also reduces the predictive bias in the original whitening filter. In other words, the chirp factor plays also a vital role in controlling the bias in the whitening process. For the Decimation–Whitening (DW) filter proposed in this paper, due to the removal of the predictive bias, the chirp factor can be designed merely for controlling whitening effect

The design of the decimation LP filter is not new in AR modeling or maximum-entropy spectral estimation. In the literature, there have been researches of the advantages of the complex decimation LP filter over the real LP filter on the improvement of the sinusoidal phase issues or neighboring frequency resolution. Especially, since the expanding of the frequency scale by two can reduce the interference at one spectral peak caused by other neighboring frequency components, a higher frequency resolution for LP estimation can be achieved. The decimation filter has been suggested in [7]–[11]. However, these alternative complex filters require computational overhead when compared with the real ones in these scenarios. In SBR, the DW filter not only has advantages but also saves half the computational complexity for evaluating LP coefficients thanks to the data reduction from decimation.

B. Examples and Comparisons

According to the standard [2], the LP in SBR should be implemented via the second-order covariance method covering 32 samples for each CEMFB subband per audio frame. Fig. 7 compares the original whitening method in SBR with the proposed method. In the figure the 32-point DFT magnitude spectra of the original CEMFB subband and the whitened ones by the original and proposed methods are depicted in the decibel (dB) domain. As can be seen in Fig. 7(a) where the subband is generated from a real white noise, the proposed method slightly alters the original spectrum, while the original method not only alters the positive spectrum but also amplifies the negative spectrum. The evaluated NSR values in this case are 0.38 and 0.93, respectively, by the original and proposed methods; the original method gives a poor NSR estimation. For the second instance in Fig. 7(b) where the subband contains a very strong sinusoid component, both methods have good whitening effect, but the proposed method results in a flatter whitened spectrum. In Fig. 7(c), where the original subband has three sinusoid components located in the frequency interval between 0 and $3\pi/2$, the original method slightly attenuates the largest one but amplifies the others. This phenomenon illustrates the interference among the components. Oppositely, the proposed method destroys the largest one without amplifying the others due to frequency scaling. Fig. 8 illustrates the better whitening effect of the proposed filter on a tone-rich signal. In Fig. 8(b) and (c), the LF decoded AAC signal is filtered by the original and proposed filters, respectively. Both chirp factor values for the filters are equal to 1, and no additional noise is added. From the HF spectra, we can see that the original filter cannot "whiten" the tonal structure, while the DW filter does better.

Another noticeable feature is that the proposed method keeps better the energy of HF than the original method. In the SBR en-

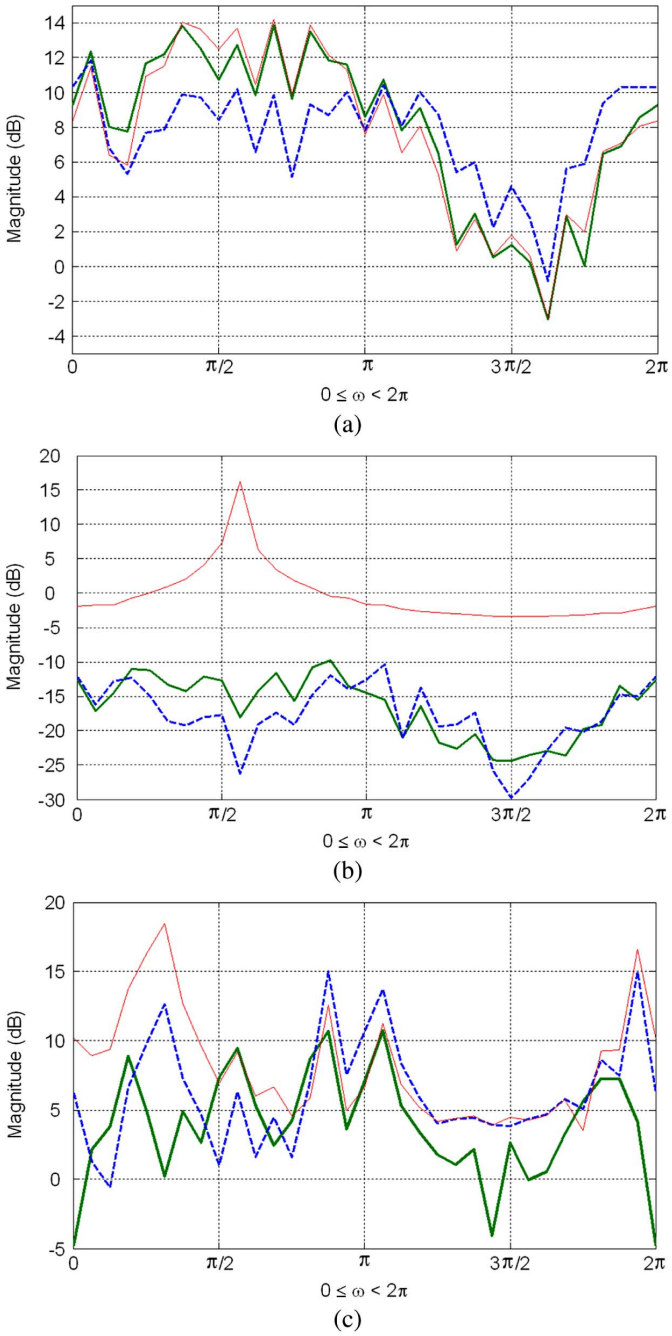

Fig. 7. Whitening comparison for the original method and the proposed method. The magnitude spectra are evaluated through 32-point DFT; thin line depicts the original signals, thick line depicts the whitened signals by the proposed method, and dash line depicts the whitened signals by the original method.

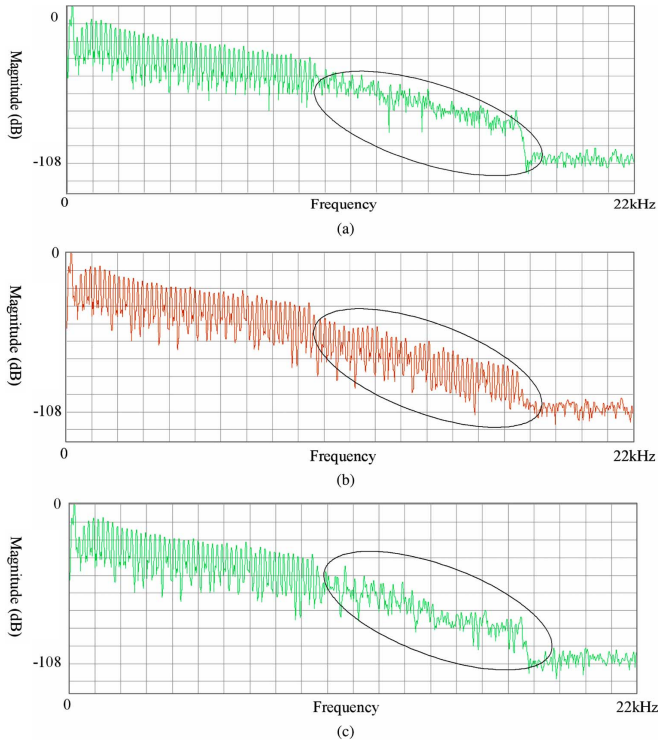

Fig. 8. Whitening comparison for the original method and the proposed method. (a) The original DFT magnitude spectrum. (b) The decoded DFT magnitude spectrum with the original whitening filter. (c) The decoded DFT magnitude spectrum with the decimation-whitening filter. The spectra are depicted in the dB domain. For both filters, the chirp factor takes 1. No additional noise is added, and the audio sampling rate is 44.1 kHz.

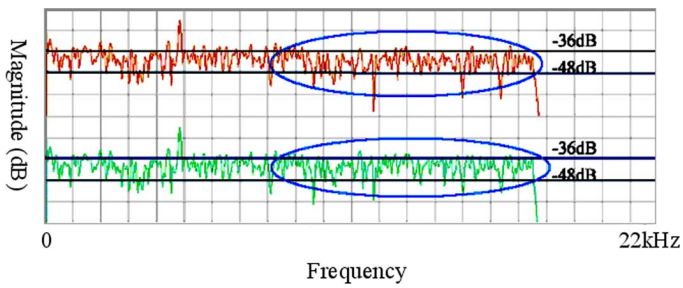

Fig. 9. Energy loss effect of the original whitening method. The depicted spectra are the decoded spectra with the original method (the upper) and the proposed method (the below), respectively. For both filters, the chirp factor takes 0.98. No additional noise is added, and the audio sampling rate is 44.1 kHz. The HF envelope of the decoded spectrum with the proposed method fits −36 dB, while that with the original method is under −36 dB.

coder, the energy of HF is calculated and recorded based on HF CEMFB subbands which are highly analytic-signal-like. Subsequently, the SBR decoder adjusts the energy of the whitened LF subbands to fit the recorded HF energy. However, as noticed in the previous discussion, the original filter has more energy leakage due to the amplification in the negative side band. The negative spectra will be filtered out by the synthesis filterbank, leading to energy loss since the energies in the negative side bands of the regenerated HF subbands have contributed to the energy estimation. The proposed DW filter has better control due to less leakage from the negative frequency range. Fig. 9 illustrates the better envelope maintenance by comparing the spectra from the two methods, where the chirp factors are 0.98 for all the replicated subbands and no additional noise is added for HF. The original signal consists of white noise and a single tone in LF.

Figs. 10 and 11 compare the performance of the DW and original LP filters on a CEMFB subband signal of an impulse in combination with four main nonzero chirp factors used in SBR. As shown in Fig. 10, the original second-order LP filter shapes two spectral valleys on the CEMFB subband signal even when the chirp factor is 0.75. When the chirp factor equals 0.6, the spectrum of the CEMFB subband is flat, but the amplification

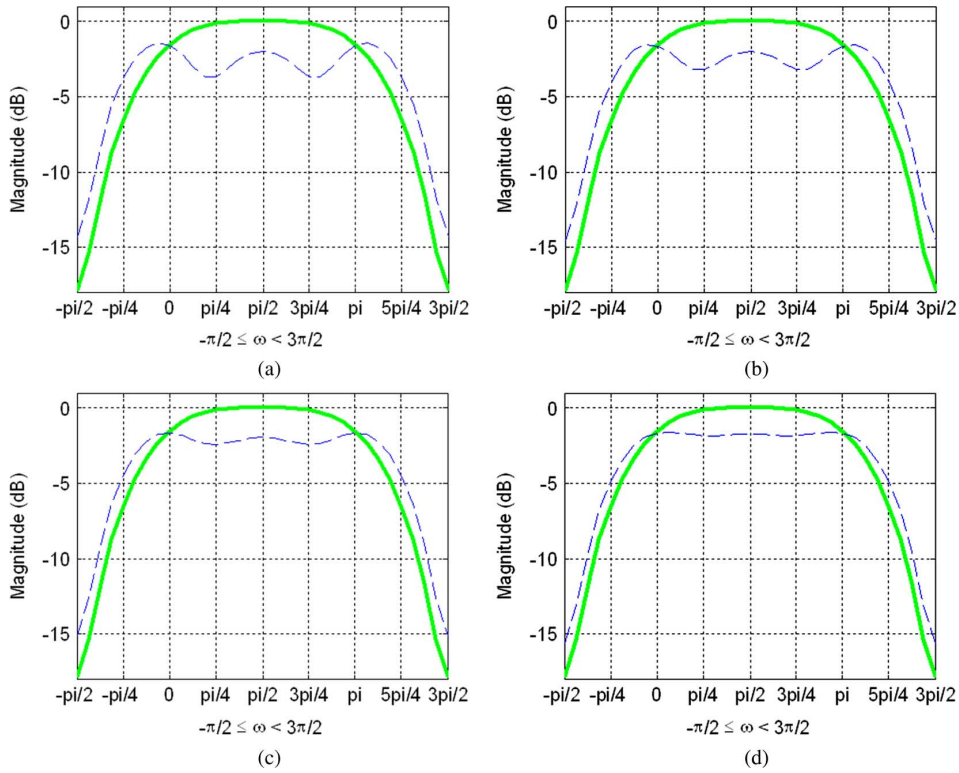

Fig. 10. Whitening processing of the original SBR second-order LP filter on a 32-point CEMFB subband signal of an impulse with different chirp factors. (a) Chirp factor equals 0.98. (b) Chirp factor equals 0.9. (c) Chirp factor equals 0.75. (d) Chirp factor equals 0.6. (Solid line depicts original signals; dashed line depicts whitened signals, where $\mathrm{pi} = \pi$).

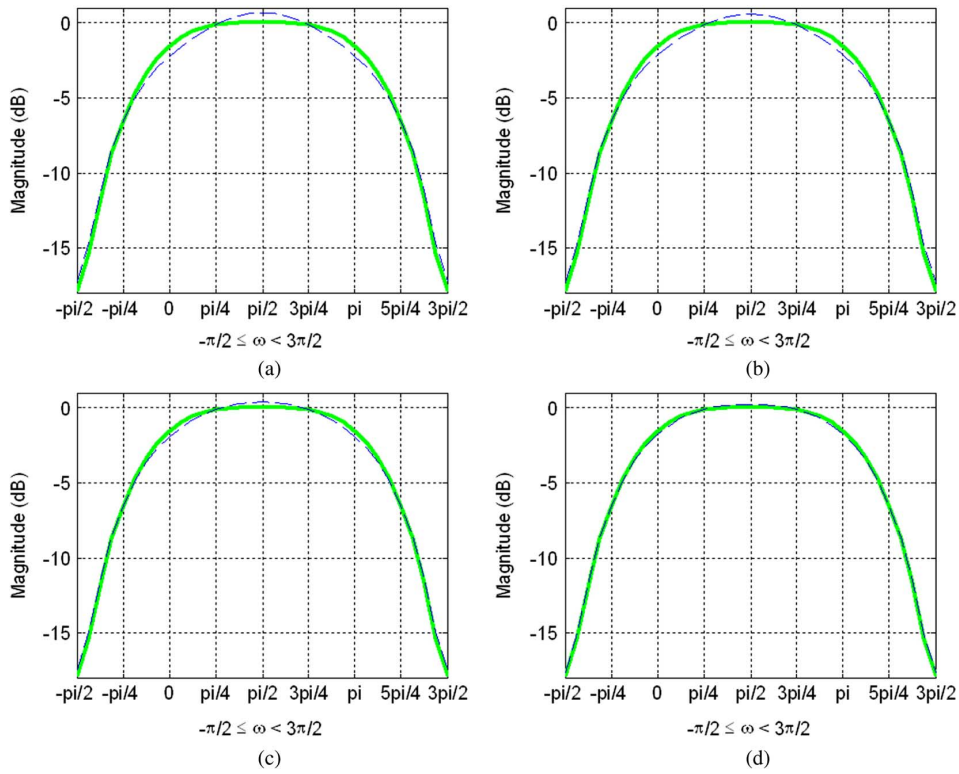

Fig. 11. Whitening processing of the DW LP filter on a 32-point CEMFB subband signal of an impulse with different chirp factors. (a) chirp factor equals 0.98. (b) Chirp $\mathrm{factor} = 0.9$. (c) Chirp $\mathrm{factor} = 0.75$. (d) Chirp $\mathrm{factor} = 0.6$. (Solid line depicts original signals; dashed line depicts whitened signals, where $\mathrm{pi} = \pi$)

on the negative side band will result in energy leakage. Comparatively, the DW filter (see Fig. 11) only slightly alters the CEMFB subband even when the chirp factor is 0.98.

C. Subjective Test

To confirm possible side effects or conflicts with other audio decoder modules, the subjective test was conducted for the 12

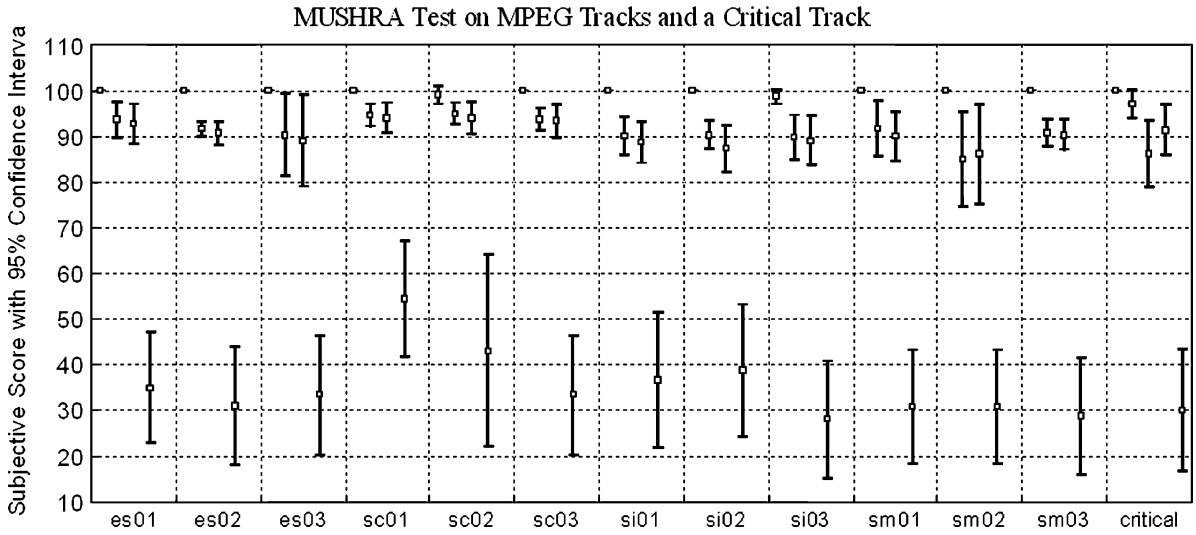

Fig. 12. Mean subjective scores of six listeners with 95% confidence intervals. For each MPEG track, the subjective scores are depicted in the order: "the hidden reference," "the decoded track with the DW filter," "the decoded track with the original filter," and "the hidden anchor"; for the critical track, the subjective scores for the critical track are depicted in the order: "the hidden reference," "the decoded track with the DW filter for chirp factor 0.98," "the decoded track with the original filter for chirp factor 0.98," "the decoded track with the original filter for chirp factor 0.75," and "the hidden anchor." The bitrate of the HE-AAC encoder takes 48 kbps.

TABLE I
MPEG TEST TRACKS (THE ORIGINAL VERSION IS 16 BITS/44.1 k SAMPLING RATE; THE COMPRESSED VERSION IS ENCODED BY THE HE-AAC ENCODER WITH BITRATE 48 kbps)

Track		Description	Duration
1	es01	vocal (Suzan Vega)	10 sec
2	es02	German speech	8 sec
3	es03	English speech	7 sec
4	sc01	Trumpet solo and orchestra	10 sec
5	sc02	Orchestral piece	12 sec
6	sc03	Contemporary pop music	11 sec
7	si01	Harpsichord	7 sec
8	si02	Castanets	7 sec
9	si03	pitch pipe	27 sec
10	sm01	Bagpipes	11 sec
11	sm02	Glockenspiel	10 sec
12	sm03	Plucked strings	13 sec

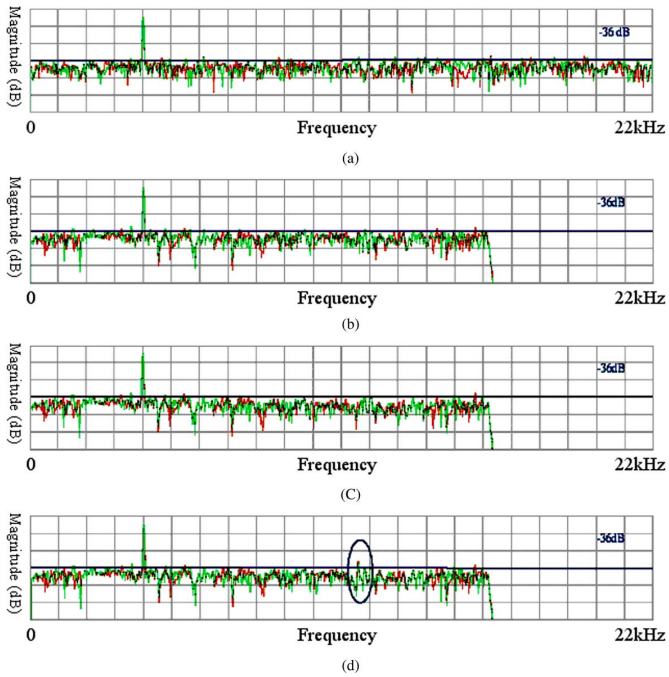

Fig. 13. Comparison of decoded spectra for the critical track. (a) The original reference signal. (b) The decoded spectrum with the DW filter for chirp factor 0.98. (c) The decoded spectrum with the original filter for chirp factor 0.98. (d) The decoded track with the original filter for chirp factor 0.75.

MPEG test tracks (see Table I) and a critical track. The test followed the Multiple Stimulus with Hidden Reference and Anchors (MUSHRA) methodology, which is recommended in ITU-R Recommendation BS.1534-1 [12] for the subjective assessment of intermediate quality level of coding systems. A total of six subjects participated in the test. In each test trial, the subjects were presented with a reference (original) signal, a hidden version of the reference signal, a hidden anchor and a certain number of test tracks. The anchor used is the 3.5-kHz low-pass version of the reference signal. The subjects were asked to score each test item by using a 100-point grading scale divided in five equal intervals labeled "excellent," "good," "fair," "poor," and "bad."

The test results are displayed in Fig. 12 as mean scores with 95% confidence interval. The intersection of the two confidence intervals related to the DW and original filters illustrates that the DW filter does not bring artifacts or conflict with other audio decoder modules in comparison to the original filter. Especially, it can be seen that the DW filter provides a better mean quality for all but one.

On the other hand, a critical track consisting of a single tone of 4 kHz and a white noise (see Fig. 13) is artificially

constructed for verifying the energy leakage issue. The duration of the critical track is 10 s. Comparing Fig. 13(c) with (d) shows that to suppress the duplicated tone, the chirp factor should take a high value (e.g., 0.98) for the original filter. It can be observed from Fig. 13(b) and (c) that with the usage of the high chirp factor 0.98, the original filter results in energy decay on the reconstructed noise floor, whereas the DW filter can keep a better noise floor. In Fig. 12, the result that with the chirp factor 0.98, the subjective quality of the DW filter is better than that of the original filter may be due to the energy leakage from the original filter. When using the smaller chirp factor 0.75, the original filter can have an improved perceptual quality but is still worse than the DW filter (with the chirp factor 0.98) in mean score.

V. Conclusion

SBR adopts a second-order LP for whitening filtering. Through modeling CEMFB subbands as analytic signals, this paper has demonstrated the predictive bias for the whitening filter. The bias increases the interference of noise components to sinusoid components in LP and leads to spectral hollows in whitened white-spectral or noise-like subbands. For removing the bias, this paper has proposed a novel filter, named the DW filter. The DW filter cannot only eliminate the bias but also reduce the interferences from sinusoid components by frequency stretching. Compared with the original filter adopted in SBR, the DW filter has three distinct features. The first is on the more accurate NSR measure that can be provided for the tonality control algorithm in the SBR encoder. The second is the less energy loss from the filtering out of the negative frequency after filterbank synthesis. The third is the reduction of computational complexity by half for the LP coefficients calculation due to the sample number reduction from the decimation.

Appendix A

First, consider the evaluation of integration K

$$K = \int_0^{\pi} \frac{1}{1 - 2r_0\cos(\omega - \theta_0) + r_0^2} d\omega = \int_{-\theta_0}^{\pi-\theta_0} \frac{1}{1 - 2r_0\cos(\omega) + r_0^2} d\omega. \quad \text{(A.1)}$$

Note that we assume $0 < r_0 < 1$ and $0 \le \theta_0 \le \pi$. We might write

$$K = \int_{\Omega} \frac{1}{1 - 2r_0 \cdot \frac{1}{2}\left(z + \frac{1}{z}\right) + r_0^2} \frac{dz}{jz} = \frac{j}{r_0^2 - 1}\left(\int_{\Omega} \frac{1}{z - r_0} dz - \int_{\Omega} \frac{1}{z - r_0^{-1}} dz\right) \quad \text{(A.2)}$$

where path Ω is the upper arc of the unit circle from $e^{-j\theta_0}$ to $e^{j(\pi-\theta_0)}$. Since K is real, (A.2) can be rewritten as

$$K = \frac{1}{1 - r_0^2}\left[\mathrm{Im}\left(\int_{\Omega} \frac{1}{z - r_0} dz\right) - \mathrm{Im}\left(\int_{\Omega} \frac{1}{z - r_0^{-1}} dz\right)\right]. \quad \text{(A.3)}$$

Using the formula $\log(z - r_0) = \log|z - r_0| + j\arg(z - r_0)$, the first integration can be evaluated as

$$\mathrm{Im}\int_{\Omega} \frac{1}{z - r_0} dz = \arg[\cos(\pi - \theta_0) - r_0 + j\sin(\pi - \theta_0)] - \arg[\cos(-\theta_0) - r_0 + j\sin(-\theta_0)]. \quad \text{(A.4)}$$

In (A.4), we can choose the branch with $-\pi < \arg(z - r_0) < \pi$ such that $\log(z - r_0)$ is analytic in the domain $\{z \in \boldsymbol{C} - \{r_0\} | -\pi < \arg(z - r_0) < \pi\}$ containing Ω. Then, in terms of the arctangent function arctan that is with range of $(-\pi/2, \pi/2)$, the integration in (A.4) can be rewritten as T defined in (19). Similarly, we can choose the branch with $0 < \arg(z - r_0^{-1}) < 2\pi$ such that $\log(z - r_0^{-1})$ is analytic in the domain $\{z \in \boldsymbol{C} - \{r_0^{-1}\} | 0 < \arg(z - r_0^{-1}) < 2\pi\}$ containing Ω. Subsequently, we have

$$\mathrm{Im}\int_{\Omega} \frac{1}{z - r_0^{-1}} dz = -\arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0) + r_0^{-1}}\right) + \arctan\left(\frac{\sin(\theta_0)}{\cos(\theta_0) - r_0^{-1}}\right). \quad \text{(A.5)}$$

Substituting the two closed forms in (19) and (A.5) into (A.3) yields (18). On the other hand, from the integral identity $\int_0^{\pi} (1 - 2r_0\cos(\omega - \theta_0) + r_0^2)/(1 - 2r_0\cos(\omega - \theta_0) + r_0^2) d\omega = \pi$, we can evaluate C as (17). Equation (16) can be derived by the technique of changing valuables.

Appendix B

To find the MMSE solution of (21), from the geometric symmetry of the solution, we might assume that $\tilde{r}_1 = \tilde{r}_2 = r, \tilde{\theta}_1 = \theta, \tilde{\theta}_2 = \pi - \theta$ and $\theta \in [0, \pi/2]$. Then we can evaluate the integration in (21) as

$$(1 + 2r^2 - 2r^2\cos(2\theta) + r^4)\pi - 8(r + r^3)\sin(\theta). \quad \text{(B.1)}$$

Deriving $\partial F/\partial\theta = 0$ and $\partial F/\partial r = 0$ yields, respectively,

$$r[1 + r^2 - \cos(2\theta)]\pi - 2(1 + 3r^2)\sin(\theta) = 0, \quad \text{(B.2)}$$

$$1 + r^2 = r\pi\sin(\theta). \quad \text{(B.3)}$$

Repeatedly substituting (B.3) into (B.2) to reduce the power of term r from 2 to 1 and using the trigonometric property $2\sin^2(\theta) + \cos(2\theta) = 1$ can give

$$r = \frac{\pi}{(\pi^2 - 4)\sin(\theta)}. \quad \text{(B.4)}$$

We can obtain θ by substituting (B.4) into (B.3) as

$$\sin(\theta) = \frac{\pi}{2\sqrt{\pi^2 - 4}}. \tag{B.5}$$

Substituting (B.5) into (B.4) yields (22). Similarly, by repeatedly substituting (B.2) and using the trigonometric property $2\sin^2(\theta) + \cos(2\theta) = 1$, we can derive (B.1) as

$$F(r,\theta) = r^2\pi[\pi^2 \sin^2(\theta) - 2\cos(2\theta)] - 8[r^2\pi \sin^2(\theta)]. \tag{B.6}$$

By substituting (B.5) and (22) into (B.6), we can obtain that the MMSE equals $\pi(\pi^2 - 8)/(\pi^2 - 4)$.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their comments that significantly improved this paper.

REFERENCES

[1] P. Ekstrand, "Bandwidth extension of audio signals by spectral band replication," in *Proc. 1st IEEE Benelux Workshop Model Based Process. Coding Audio*, Leuven, Belgium, Nov. 2002, pp. 53–58.

[2] Bandwidth Extension Pattaya, Thailand, ISO/IEC JTC1/SC29/WG11/N5570, ISO/IEC, 14496-3:2001/FDAM1, Mar. 2003.

[3] M. Wolters, K. Kjörling, D. Homm, and H. Purnhagen, "A closer look into MPEG-4 high efficiency AAC," in *Proc. AES 115th Conv.*, New York, Oct. 2003, preprint 5871.

[4] C. M. Liu, H. W. Hsu, and W. C. Lee, "Compression artifacts in perceptual audio coding," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 681–695, May 2008.

[5] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[6] L. S. Marple, "Computing the discrete-time 'analytic' signal via FFT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2600–2603, Sep. 1999.

[7] S. M. Kay, "Maximum entropy spectral estimation using the analytic signal," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 5, pp. 467–469, Oct. 1978.

[8] L. B. Jackson, D. W. Tufts, F. K. Soong, and R. M. Rao, "Frequency estimation by linear prediction," in *Proc. IEEE ICASSP*, 1978, pp. 332–356.

[9] S. M. Kay, "Fourier-autoregressive spectral estimation," in *Proc. IEEE ICASSP*, 1979, pp. 162–165.

[10] M. P. Quirk and B. Liu, "Improving resolution for autoregressive spectral estimation by decimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 3, pp. 630–637, Jun. 1983.

[11] T. Shimamura, N. Sakaguchi, and S. Takahashi, "Frequency estimation using the analytic signals by decimation," in *Proc. IEEE Pacific Rim Conf. Commun., Comput., Signal Process. (PACRIM 1989)*, Victoria, BC, Canada, Jun. 1–2, 1989, pp. 540–543.

[12] "Recommendation B.S. 1543-1: Method for the subjective assessment of intermediate sound quality (MUSHRA)," Int. Telecomm. Union, Geneva, Switzerland, 2001, ITU-R.

Han-Wen Hsu was born in Tainan, Taiwan, in October 1977. He received the B.S. degree from the Division of Applied Mathematics, Department of Mathematics, National Tsing Hua University, Hsinchu, Taiwan, in 2000 and the M.S. and Ph.D. degrees from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, in 2004 and 2010, respectively.

His research interests are in audio coding and signal processing.

Chi-Min Liu received the M.S. and Ph.D. degrees in electronics from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1987 and 1991, respectively.

He is currently a Professor in the Department of Computer Science, NCTU. His research interests include audio compression, audio effects, and video compression. He leads the Perceptual Signal Processing Laboratory at NCTU.