Bao *et al. EURASIP Journal on Audio, Speech, and Music Processing* 2014, **2014**:1
http://asmp.eurasipjournals.com/content/2014/1/1

EURASIP Journal on
Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH **Open Access**

A blind bandwidth extension method for audio signals based on phase space reconstruction

Chang-Chun Bao*, Xin Liu, Yong-Tao Sha and Xing-Tao Zhang

Abstract

Bandwidth extension is an effective technique for enhancing the quality of audio signals by reconstructing their high-frequency components. In this paper, a novel blind bandwidth extension method is proposed based on phase space reconstruction. Phase space reconstruction is introduced to convert the low-frequency modified discrete cosine transform coefficients of wideband audio to a multi-dimensional space, and the high-frequency modified discrete cosine transform coefficients of the audio signal are reconstructed by a non-linear prediction model. The performance of the proposed method was evaluated through objective and subjective tests. It is found that the proposed method achieves a better performance than the typical linear extrapolation method, and its performance is comparable to the conventional efficient high-frequency bandwidth extension method.

Keywords: Audio coding, Bandwidth extension, High-frequency reconstruction, Phase space reconstruction

1 Introduction

According to formal and informal listening tests, the listeners usually prefer the bandwidth-limited audio signals over the heavily distorted full-band signals. So, the high-frequency components (HFC) of audio signals are partly or completely discarded in many audio coding methods at low bit rates in order to increase coding efficiency. If a high-frequency reconstruction module is embedded into audio codecs, the quality of the reproduced audio signals could be improved.

Bandwidth extension (BWE) plays an important role in high-frequency reconstruction of audio signals. Generally, BWE can be divided into non-blind BWE and blind BWE. In the non-blind BWE method, some side information, such as time/frequency envelope of high-frequency (HF) band, noise floor, level of inverse filtering, and additional sine signals, should be extracted and coded in advance at the encoder and transmitted together with the encoded low-frequency components (LFC). At the decoder, the components in the low-frequency (LF) band were copied into the HF band based on the received side information [1]. This method is often called spectral band replication (SBR) [2,3]. A well-known audio codec utilizing SBR is high-efficiency advanced audio coding (HE-AAC) [4,5], which has been applied in mobile multimedia players, mobile phones, and digital radio services. In addition, the non-blind BWE methods used in extended adaptive multi-rate - wideband (AMR-WB+) [6] and audio/video coding standard for mobile multimedia applications (AVS-M) [7] also exhibit good BWE performance by adopting a linear prediction model to describe the spectral envelope and extending the bandwidth of audio signals based on the time/frequency envelope. Compared to LFC coding, non-blind BWE can help an audio codec to produce a much higher decoding quality with side information, but it takes up more channel resource for transmission. Thus, the non-blind BWE method is unusable when the transmission channel cannot afford enough bit rates for additional side information.

For the blind BWE, the HFC of audio signals are often completely discarded at the encoder, and there is no side information related to the HFC for coding and transmission. Only the LFC are coded at the encoder. The high-frequency reconstruction completely depends on the decoded LFC. Blind BWE can be easily applied within different audio codecs. Traditional blind BWE methods have been studied extensively for narrowband speech based on the speech generation model [1]. However, there is as yet no efficient method for blind BWE of audio signals. A straightforward method is to linearly extrapolate the HFC

* Correspondence: baochch@bjut.edu.cn
Speech and Audio Signal Processing Lab, School of Electronic Information and Control Engineering, Beijing University of Technology, 100124, Beijing, China

from the LFC under the assumption that the audio amplitude spectrum with logarithm scale is linearly declined with the frequency increase [8,9]. However, this assumption of audio amplitude spectrum is not true in most cases due to the complicated characteristics of the audio spectrum. Actually, audio signals have more non-linear characteristics, and an efficient non-linear prediction method is essential for reconstructing the HFC in blind BWE. Efficient high-frequency bandwidth extension (EHBE) [10] could reproduce some new HF harmonic components using non-linear filtering methods. In the EHBE method, after the highest octave presented in the decoded LF information was extracted as the fundamental by a band-pass filter, the harmonics could be created by half-wave rectification. Then, the desired part of the complete harmonic signal was extracted by another band-pass filter and scaled by gain *G*. Combined with the delayed input signals, the full-frequency audio signals could be obtained. Due to frequency mixing, auditory distortion is also perceived following the application of the EHBE method.

In this paper, a blind high-frequency reconstruction method of audio signals based on phase space reconstruction (PSR) and non-linear prediction is proposed. Here, PSR is used to convert the LF modified discrete cosine transform (MDCT) coefficients of wideband audio to a multi-dimensional space. In order to improve the performance, the energy and harmonic components of the reconstructed HF spectrum are further adjusted according to listening perception. The objective and subjective evaluations show that the proposed method achieves a better performance than linear extrapolation (LE) method [8] and is comparable to the EHBE method [10].

The outline of this paper is as follows: First, we describe the PSR of audio signals and discuss the calculation of embedding dimension and embedding delay related to PSR. Section 3 describes the high-frequency reconstruction principle of audio signals, followed by a more detailed description of the prediction of the HF-MDCT coefficients and the adjustment of energy and harmonics of HFC. The performance of the proposed method is evaluated in Section 4. Finally, the conclusions are given in Section 5.

2 Phase space reconstruction of audio signals

Motivated by the previous works [11,12], the PSR method, which is similar to the delay reconstruction method [13], is used to convert the LF-MDCT coefficients into a multi-dimensional space. A non-linear prediction model is built up in the phase space to simulate the hidden relationship between the given phase points and the unknown MDCT coefficients. By using this model, we can restore the audio spectrum of the HF components from the LF components in the phase space. So, the primary problem is how to build a multi-dimensional phase space from a one-dimensional audio signal.

The one-dimensional vector **x** derived from the LF-MDCT coefficients is represented as

$$\mathbf{x} = \{x(k), k = 1, 2, ..., M\}, \tag{1}$$

where k and M are the index and the number of LF-MDCT coefficients, respectively. The phase space of the one-dimensional vector **x** can be reconstructed by choosing a proper embedding dimension m and an embedding delay τ according to the similar principle from the delay reconstruction method [13]. Through PSR, the M LF-MDCT coefficients are mapped into M-$(m$-$1)\tau$ phase points, where the dimension number of each phase point is m. All the phase points compose a phase space **Y**, and it can be represented by a $[M$-$(m$-$1)\tau] \times m$ matrix,

$$\mathbf{Y}=\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{M-(m-1)\tau} \end{bmatrix}$$

$$= \begin{bmatrix} x(1) & x(1+\tau) & \cdots & x(1+(m-1)\tau) \\ x(2) & x(2+\tau) & \cdots & x(2+(m-1)\tau) \\ \vdots & \vdots & & \vdots \\ x(M-(m-1)\tau) & x(M-(m-2)\tau) & \cdots & x(M) \end{bmatrix}, \tag{2}$$

where $\mathbf{y}_k$, $k = 1,2,..., M$-$(m$-$1)\tau$ is a m-dimensional row vector, and it represents a phase point in phase space, the embedding dimension m represents the minimum dimension number of phase space, and the embedding delay τ describes the distance between adjacent components of each phase point. Thus, once the phase space of LF-MDCT coefficients is reconstructed, the relationship between the phase points and MDCT coefficients can be described by a non-linear model. The HFC of audio signals can be intra-frame restored from these phase points in phase space based on non-linear prediction, according to the assumption of the strong correlation between HF and LF spectra of the audio signals.

Here, we will give a sample to explain the reconstruction procedure of HF-MDCT coefficients. We assume that $M = 6$, $m = 3$, $\tau = 1$, and the reconstructed phase space including four phase points $\mathbf{y_1}$, $\mathbf{y_2}$, $\mathbf{y_3}$, and $\mathbf{y_4}$ can be represented as follows:

$$\mathbf{Y} = \left[\begin{array}{c} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \\ \hline \mathbf{y}_5 \\ \mathbf{y}_6 \\ \vdots \end{array}\right] = \left[\begin{array}{ccc} x(1) & x(2) & x(3) \\ x(2) & x(3) & x(4) \\ x(3) & x(4) & x(5) \\ x(4) & x(5) & x(6) \\ \hline x(5) & x(6) & x(7) \\ x(6) & x(7) & x(8) \\ \vdots & \vdots & \vdots \end{array}\right] \tag{3}$$

According to the PSR principle, we can obtain the phase points $\mathbf{y}_5$, $\mathbf{y}_6$,..., containing the HF-MDCT coefficients, $x(7)$, $x(8)$,.... The (linear or non-linear) relationship between phase point $\mathbf{y}_5$ and its adjacent phase points, such as $\mathbf{y}_3$ and $\mathbf{y}_4$, can be used to predict the component $x(7)$ included in phase point $\mathbf{y}_5$. Once the component $x(7)$ within the phase point $\mathbf{y}_5$ is determined, we can use a similar method to estimate the component $x(8)$ within phase point $\mathbf{y}_6$, and so on. This procedure does not stop until the final component corresponding to the cutoff frequency of the full band is determined.

In order to reconstruct the phase space, two important parameters m and τ should be calculated in advance. The details of calculating embedding dimension m and embedding delay τ will be given in the following two sub-sections.

2.1 Calculation of embedding delay τ

The embedding delay τ should be neither too small nor too large. If τ is too small, the difference between the adjacent components of each phase point in the phase space will become too small. If τ is too large, there will be no any correlation between adjacent components of each phase point in the phase space. These two cases will not correctly reflect the true relationship between the MDCT coefficients, and the PSR will be ineffective. We found that for the MDCT spectrum of audio signals, a slight change of τ will lead to the adjacent phase points becoming completely uncorrelated. In this case, BWE from LF to HF is impossible. Therefore, a trade-off of embedding delay τ is crucial for the PSR.

In this paper, we use the de-biasing autocorrelation function (DB-ACF) method [13] which has a lower computational complexity to calculate the embedding delay τ. The embedding delay τ obtained by this method not only maintains a lower correlation but also has a non-complete independence between adjacent components of each phase point in the phase space. The de-biasing autocorrelation function $R(i)$ of LF-MDCT coefficients is given by

$$R(i) = \frac{1}{M-i}\sum_{k=1}^{M-i}\{[x(k)-\bar{x}][x(k+i)-\bar{x}]\} \tag{4}$$

where $\bar{x}$ denotes the mean of LF-MDCT coefficients, and $i = 0,1,2, \ldots, M - 1$ is the index of $R(i)$.

It is found that the phase space will be reconstructed as long as the embedding delay τ is limited to a certain range, even though τ has a small deviation. Here, when the first zero value or the first minimum value of $R(i)$ appears, the corresponding index i is chosen as the embedding delay τ. This can reduce the computational complexity since it does not need to compute all $R(i)$.

2.2 Calculation of embedding dimension m

In this paper, we adopt the false nearest neighbors (FNN) [14] method to calculate the embedding dimension m. The basic idea is to gradually determine the false nearest neighbors of each phase point by increasing the dimension number of phase space. The dimension number that makes the adjacent phase points completely unfolded is chosen as the final embedding dimension m.

For a d-dimensional phase space, the arbitrary phase point $\mathbf{y}_k$ has a nearest neighbor $\mathbf{y}_k^{\mathrm{NN}} = \{x^{\mathrm{NN}}(k), x^{\mathrm{NN}}(k + \tau), x^{\mathrm{NN}}(k + 2\tau), \ldots , x^{\mathrm{NN}}(k + (d - 1)\tau)\}$. Let the Euclidean distance between phase points $\mathbf{y}_k$ and $\mathbf{y}_k^{\mathrm{NN}}$ be $D_d(k)$. When the dimension number of phase space is increased to $d + 1$ from d, the Euclidean distance between phase points $\mathbf{y}_k$ and $\mathbf{y}_k^{\mathrm{NN}}$ is denoted by $D_{d+1}(k)$. The relationship between $D_{d+1}(k)$ and $D_d(k)$ is given by [14]

$$D^2{}_{d+1}(k) = D^2{}_d(k) + \left|x(k+\tau d)-x^{\mathrm{NN}}(k+\tau d)\right|^2. \tag{5}$$

To verify whether the nearest neighbor of each phase point is false or not, the relative change of Euclidean distance between phase point and its nearest neighbor is denoted as $f_d(k)$ when the dimension number of phase space is changed to $d + 1$ from d, i.e., $f_d(k)$ is defined by [14]

$$\begin{aligned} f_d(k) &= \left[\frac{D^2{}_{d+1}(k)-D^2{}_d(k)}{D^2{}_d(k)}\right]^{\frac{1}{2}} \\ &= \frac{\left|x(k+\tau d)-x^{\mathrm{NN}}(k+\tau d)\right|}{D_d(k)}. \end{aligned} \tag{6}$$

If $f_d(k) > f_{\mathrm{D}}$, the phase point $\mathbf{y}_k^{\mathrm{NN}}$ is determined as the false nearest neighbor of the phase point $\mathbf{y}_k$, where f_{D} is a threshold value used for verifying whether the nearest neighbor of phase point is false or not. Here, f_{D} is a fixed value between 10 and 50.

In order to find the embedding dimension m, the Euclidean distance $D_d(k)$ and $D_{d+1}(k)$ between each phase point and its nearest neighbors are first calculated by increasing the dimension number of phase space d with step length 1 given the embedding delay τ. Then, $f_d(k)$ is obtained by Equation (6). By comparing $f_d(k)$ with f_D, the false nearest neighbors for all the phase points are verified. Furthermore, the percentage of the false nearest neighbors in all nearest neighbors of all phase points is defined as $\beta(d)$.

In our experiment, the initial value of d is set to 1. If $\beta(d) > \beta_D$, where β_D is a percentage threshold between 5% and 10%, d is increased by 1. The comparison between $\beta(d)$ and β_D will not stop until $\beta(d) < \beta_D$. The value d that makes $\beta(d)$ to be less than β_D is chosen as the final embedding dimension m.

Once the embedding delay τ and the embedding dimension m are determined, the phase space can be reconstructed from audio spectrum by Equation (2). In order to visualize the structure of audio spectrum, the phase space could be mapped into three-dimensional space. As a visualized example, we selected an audio signal produced by a violin to analyze the phase trajectories in the reconstructed phase space. The semitone of violin segment is D4, and the pitch is about 296 Hz. Embedding delay and embedding dimension were set to 3 and 1, respectively. Audio signals with the length of 20 ms were transformed by MDCT. The spectral envelope was removed to obtain the fine spectrum of audio signals, and the phase space was reconstructed. The trajectories of the fine spectrum in the phase space for LF and HF signals are demonstrated in Figures 1 and 2, respectively. It is manifest that the phase trajectories turn dispersed with the increase of frequency, but nearly all the phase vectors are localized in a certain range to form a hyper-ellipsoidal structure. Similar results have been obtained for audio spectra of other orchestral instruments, symphony, and pop music in our experiments. This indicates that audio spectrum is characterized by regular dynamic structure and is predictable. Moreover, the audio spectra of some percussion music and live background sound were also analyzed in three-dimensional phase space. Although the phase vectors of these two cases are scattered in a disorderly way over a certain area, the trajectories of both LF and HF fine spectra show similar characteristics of randomness. Based on this fact, the non-linear prediction method is introduced in this paper to recover the HF fine spectrum from the LF spectrum.

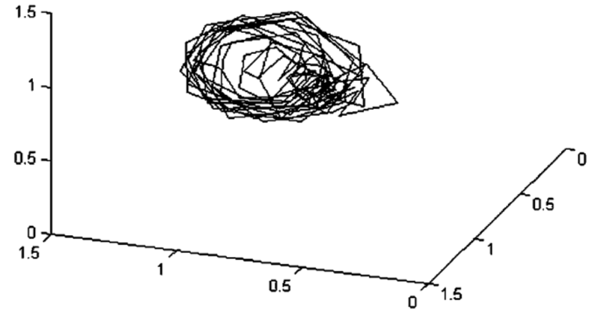


Figure 1 Example of phase trajectory for LF fine spectrum of audio signal.

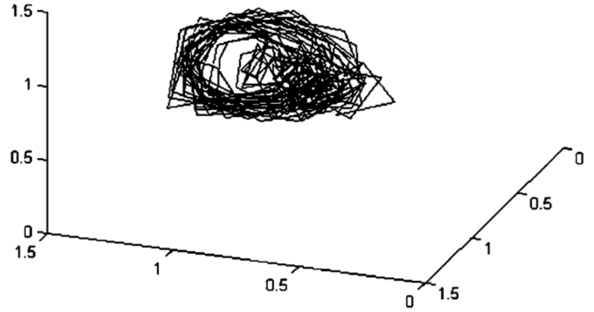


Figure 2 Example of phase trajectory for HF fine spectrum of audio signal.

3 High-frequency reconstruction of audio signals

The block diagram of the proposed high-frequency reconstruction method of audio signals is shown in Figure 3. The high-frequency reconstruction includes five steps:

Step 1. Calculate LF-MDCT coefficients of audio signals;
Step 2. Reconstruct the phase space of LF-MDCT coefficients;
Step 3. Predict the HF-MDCT coefficients;
Step 4. Adjust the energy and harmonics of HFC;
Step 5. Take an inverse MDCT of full-band audio signals.

Step 1 and step 5 are well known, so we will not discuss them here. Step 2 has been described in Section 2. In this section, we will discuss step 3 and step 4, respectively.

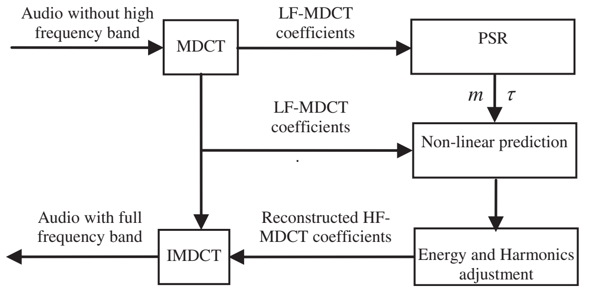

Audio without high frequency band
MDCT
LF-MDCT coefficients
PSR
m τ
LF-MDCT coefficients
Non-linear prediction
Audio with full frequency band
IMDCT
Reconstructed HF-MDCT coefficients
Energy and Harmonics adjustment

Figure 3 Block diagram of the proposed high-frequency reconstruction method.

3.1 Non-linear prediction of HF MDCT coefficients

Once the phase space is reconstructed based on LF-MDCT coefficients, the HF-MDCT coefficients can be predicted according to the hidden relationship between the given phase points and the unknown MDCT coefficients [12]. In this paper, the non-linear prediction method is utilized to obtain HF-MDCT coefficients based on the laws that the adjacent phase points have similar characteristics. We use 2τ phase points at the bottom of Equation (2) to predict the HF-MDCT coefficients because they have more relationship with HF-MDCT coefficients. The high-frequency band is divided into L regions in terms of 2τ interval, and the cutoff frequency of audio signals is located in the Lth region.

The block diagram of the non-linear prediction method of HF-MDCT coefficients is shown in Figure 4. At first, the LF phase point corresponding to the predicted HF-MDCT coefficient is determined, and its neighbors are selected from the LF phase points. Then, the weight parameters of non-linear prediction are calculated from the neighbors though the recursive least square (RLS) algorithm, according to minimize the weighted error. Finally, the HF-MDCT coefficients are recovered by using non-linear prediction point by point until the frequency reaches up to the cutoff frequency 14 kHz.

The HF-MDCT coefficients from regions 1 to L can be predicted by the following equation:

$$x(k_{\mathrm{c}}+2\tau\cdot l)=\sum_{i=0}^{m-1}w_0(i)h[x(k_{\mathrm{c}}-i\tau)]+\sum_{i-0}^{m-1}w_1(i)x(k_{\mathrm{c}}-i\tau)+C, \tag{7}$$

where $k_{\mathrm{c}} = M - 2\tau + 1,\ M - 2\tau + 2,\ \ldots,\ M$, represents the index of the highest dimension components of 2τ phase points from the LF band, $l = 1, 2, \ldots, L$ is the region index of high-frequency band, C is a constant, $w_0(i)$ is the non-linear weight coefficients, $w_1(i)$ is the linear weight coefficients, and $h(x)$ is a non-linear kernel function. Our experiments show that the following non-linear kernel function is suitable for the HF-MDCT coefficients prediction:

$$h(x)=\frac{1}{1+e^{-x}}. \tag{8}$$

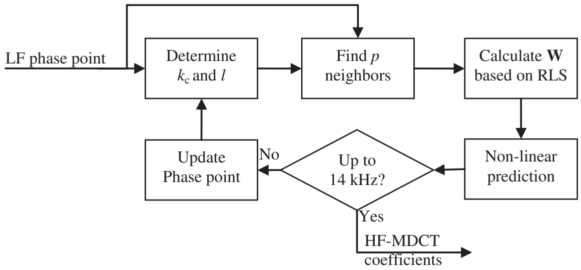


Figure 4 Block diagram of non-linear prediction of HF-MDCT coefficients.

In Equation (7), the region index l of high-frequency band starts to count from 1 and those phase points used for prediction are updated by circularly increasing k_{c}. Thus, we can use Equation (7) to predict the HF-MDCT coefficients within the L bands.

We can find p neighbors for each phase point whose highest dimension component is $x(k_{\mathrm{c}})$, $k_{\mathrm{c}} = M - 2\tau + 1$, $M - 2\tau + 2, \ldots, M$. The highest dimension components of these p neighbors are denoted as $x(k_1), x(k_2), .., x(k_p)$, respectively. By minimizing the squared error ε between these p components and the predicted value $x(k_{\mathrm{c}} + 2\tau l)$, $k_{\mathrm{c}} = M - 2\tau + 1,\ M - 2\tau + 2, \ldots, M$, $l = 1, 2, \ldots, L$ which is derived from Equation (7), the weight vector $\mathbf{W}$ composed of the weight coefficients $w_0(i)$, $w_1(i)$ and the constant C in Equation (7) can be estimated by the following cost function:

$$\varepsilon=\sum_{i=1}^{p}\gamma^{p-i}\left\{\theta_i\left[x(k_i)-\mathbf{W}^{\mathrm{T}}\mathbf{Z}\right]\right\}^2, \tag{9}$$

where p varies from $2m$ to $2m + 10$. A negative effect on prediction will occur if p exceeds this range. $\gamma = 0.9$ is a forgetting factor, $\mathbf{W} = [w_0(0),\ldots, w_0(m\text{-}1), w_1(0),\ldots, w_1(m\text{-}1), C]^{\mathrm{T}}$, $\mathbf{Z} = [h[x(k_{\mathrm{c}})],\ldots,h[x(k_{\mathrm{c}}\text{-}(m\text{-}1)\tau)], x(k_{\mathrm{c}}),\ldots, x(k_{\mathrm{c}}\text{-}(m\text{-}1)\tau),1]^{\mathrm{T}}$, θ_i is a weight used for adjusting each neighbor's effect on the prediction, that is, θ_i will be given a larger value for the component $x(k_i)$ who has a shorter distance from the prediction value $x(k_{\mathrm{c}} + 2\tau l)$. Here, θ_i is defined by

$$\theta_i=\frac{e^{-(D_i-D_{\min})}}{\sum_{j=1}^{p}e^{-\left(D_j-D_{\min}\right)}}, \tag{10}$$

where D_i, $i = 1, 2,\ldots, p$ is the Euclidean distance between $x(k_1)$, $i = 1, 2, \ldots, p$, and $x(k_{\mathrm{c}} + 2\tau l)$, $k_{\mathrm{c}} = M - 2\tau + 1$, $M - 2\tau + 2,\ldots, M$, $l = 1, 2, \ldots, L$, $D_{\min}$ is the minimum value of D_i, $i = 1, 2,\ldots, p$.

The weight vector $\mathbf{W}$ in Equation (9) can be obtained based on RLS [15]. The recursive equation is given by

$$\mathbf{W}_i=\mathbf{W}_{i-1}+\mathbf{Q}_iE_i, \tag{11}$$

where the error E_i and the gain vector $\mathbf{Q}_i$ are, respectively, defined as

$$E_i=\theta_i\left[x(k_i)-\mathbf{W}_{i-1}^{\mathrm{T}}\mathbf{Z}_i\right] \tag{12}$$

$$\mathbf{Q}_i = \frac{\mathbf{R}_{i-1}\mathbf{Z}_i}{\gamma + \mathbf{Z}_i^{\mathrm{T}}\mathbf{R}_{i-1}\mathbf{Z}_i}. \tag{13}$$

The intermediate variable $\mathbf{R}_i$ is presented as

$$\mathbf{R}_i = \frac{1}{\gamma}\left[\mathbf{R}_{i-1} = \frac{\mathbf{R}_{i-1}\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\mathbf{R}_{i-1}}{\gamma + \mathbf{Z}_i^{\mathrm{T}}\mathbf{R}_{i-1}\mathbf{Z}_i}\right]. \tag{14}$$

3.2 Energy and harmonic component adjustment of HF-MDCT coefficients

Listening experiments show that the predicted HF-MDCT coefficients may make the energy of the HF band larger than the original one. So, the HF-MDCT coefficients should be modified in order to reduce this energy change. We uniformly divide the full-frequency band into N_{sub} sub-bands. The spectral envelope $S_{\mathrm{e}}(i)$ in the *i*th sub-band is used to describe the energy of HFC and can be calculated by

$$S_{\mathrm{e}}(i) = \sqrt{\frac{1}{M_{\mathrm{sub}}}\sum_{j=0}^{M_{\mathrm{sub}}-1} x^2(M_{\mathrm{sub}}\cdot i + j)}, \tag{15}$$

where i represents the index of sub-bands, M_{sub} is the number of MDCT coefficients in the *i*th sub-band, and *j* is the index of MDCT coefficients in the *i*th sub-band.

The envelope ratio $\eta(i)$ between two adjacent sub-bands is defined by

$$\eta(i) = \frac{S_{\mathrm{e}}(i)}{S_{\mathrm{e}}(i-1)}, i = 1, 2, \cdots N_{\mathrm{sub}}-1. \tag{16}$$

The mean of all $\eta(i)$ between two adjacent low-frequency sub-bands is denoted as A_{avg}. The energy changes can be reduced by diminishing the spectral envelope. The envelope attenuation factor ξ is defined by

$$\xi = \begin{cases} A_{\mathrm{avg}}, & A_{\mathrm{avg}} < A_{\mathrm{D}} \\ A_{\mathrm{D}}, & A_{avg} \geq A_{\mathrm{D}} \end{cases}, \tag{17}$$

where A_{D} is the threshold value of ξ and here it is set to 0.95.

The final energy modification of HF band can be realized by modifying HF MDCT coefficients in each sub-band with the following equation:

$$\hat{x}(M_{\mathrm{sub}}\cdot i + j) = x(M_{\mathrm{sub}}\cdot i + j)\cdot\frac{\xi}{\eta(i)}, \tag{18}$$

where i represents the index of HF sub-bands, M_{sub} is the number of MDCT coefficients in the *i*th sub-band, and *j* is the index of MDCT coefficients in the *i*th sub-band.

We found that the harmonic components are not sufficient in the reconstructed HF spectrum. It is necessary to adjust the high-frequency harmonic components with some LF harmonic components. Our experiments have shown that the audio quality can be improved by introducing some LF harmonic components into the HF spectrum. We use the following four steps to adjust the harmonic components.

Step 1. Take fast Fourier transform (FFT) on the LF-MDCT coefficients $x(i)$, $i = 1, \ldots, M$, and the reconstructed HF-MDCT coefficients $\hat{x}(i)$, $i = M + 1, \ldots, 2M$, respectively. The amplitude and phase of discrete Fourier transform (DFT) coefficients derived from LF-MDCT coefficients and HF-MDCT coefficients are denoted as $a_{\mathrm{LF}}(f)$, $\phi_{\mathrm{LF}}(f)$, $a_{\mathrm{HF}}(f)$, $\phi_{\mathrm{HF}}(f)$, $f = 1, \ldots, M$;

Step 2. Select five DFT coefficients derived from the LF-MDCT coefficients whose amplitudes are larger than others, and these coefficients' frequency indexes are denoted as f_i, $i = 1, \ldots, 5$. Replace $a_{\mathrm{HF}}(f_i)$, $i = 1, \ldots, 5$ by $a_{\mathrm{LF}}(f_i)$ $i = 1, \ldots, 5$, and keep $\phi_{\mathrm{HF}}(f_i)$, $i = 1, \ldots, 5$ unchanged;

Step 3. Take inverse FFT on the adjusted DFT coefficients, $a_{\mathrm{HF}}(f)\exp[i\phi_{\mathrm{HF}}(f)]$;

Step 4. Modify high-frequency spectrum energy with the aforementioned method.

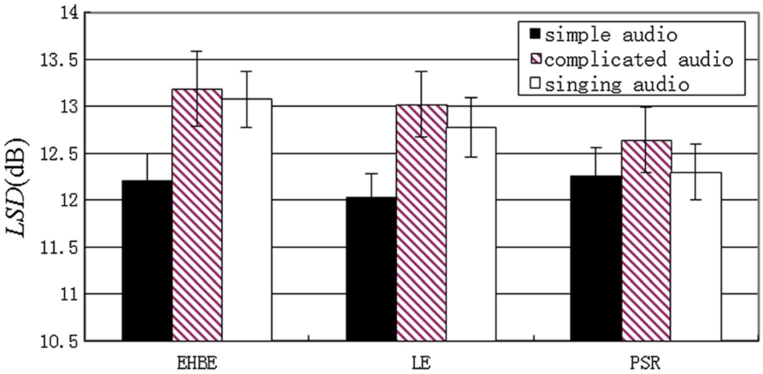
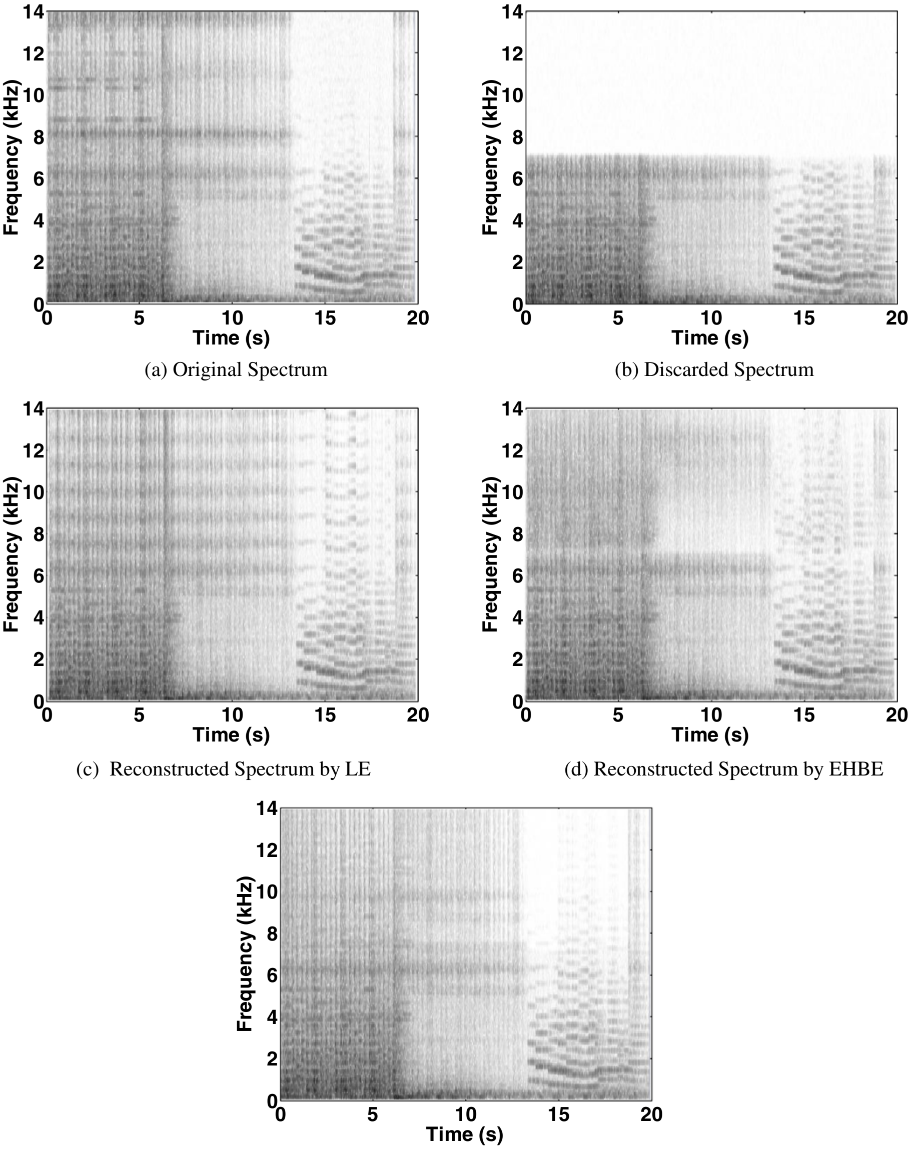


Figure 5 LSD with 95% confidence intervals for the three BWE methods.

Thus, we finish the reconstruction of the HF-MDCT coefficients. The full-band MDCT coefficients are obtained by connecting the LF-MDCT and HF-MDCT coefficients. The audio signals in the time domain can be reconstructed by an inverse MDCT.

4 Performance evaluation

In this section, the reconstructed audio signals that are independent of audio codecs are firstly used for objective evaluation. According to the definition from the International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) [16], super wideband (SWB) audio signals with a 14-kHz bandwidth were sampled at 32 kHz and were transformed into frequency domain though MDCT with a Kaiser-Bessel window. Every 20 ms, the most recent 1,280 audio samples were fed to the MDCT and were transformed into a frame of 640 spectral coefficients centered at 25-Hz intervals. First, the last 360 coefficients of these audio signals above 7 kHz were discarded to obtain the wideband (WB) audio signals with a 7-kHz bandwidth. Then, the proposed BWE method was employed to extend the bandwidth of these WB audio signals to 14 kHz. Finally, we compared the performance among the proposed method, LE method, and EHBE method. Here, the half-wave rectification was used as the non-linear device in



(a) Original Spectrum

(b) Discarded Spectrum

(c) Reconstructed Spectrum by LE

(d) Reconstructed Spectrum by EHBE

(e) Reconstructed Spectrum by PSR

Figure 6 An example of audio spectrogram reconstruction. (a) Original spectrum. **(b)** Discarded spectrum. **(c)** Reconstructed spectrum by LE. **(d)** Reconstructed spectrum by EHBE. **(e)** Reconstructed spectrum by PSR.

our experiment with respect to EHBE method, and it provides the best performance compared with other non-linear devices according to the experimental results.

In the objective evaluation, 18 audio files were chosen for testing. The length of each audio signal is between 10 to 20 s, and they were sampled at 32 kHz. These audio signals used for testing were divided into three types: simple audio, complicated audio, and singing audio. Each type includes six audio files. For simple audio signals, the number of instruments performed is less than four. For the complicated audio signals, the number of instruments performed is much larger than four. Here, singing with accompaniment is called singing audio. Moreover, the input of BWE is WB audio signals generated from SWB audio signals, and its level was normalized to −26 dBov.

The log spectral distortion (LSD) [17] was used for objective evaluation in this paper. The LSD measurement based on FFT power spectrum was given by

$$d_{\mathrm{LSD}}(i) = \sqrt{\frac{1}{N_{\mathrm{high}} - N_{\mathrm{low}} + 1} \sum_{n-N_{\mathrm{low}}}^{N_{\mathrm{high}}} \left[10 \log_{10} \frac{P_i(n)}{\hat{P}_i(n)} \right]^2}, \tag{19}$$

where i denotes the index of audio frame, and P_i and $\hat{P}_i$ are the power spectra of the original SWB audio signals and the extended audio signals, respectively. N_{high} and N_{low} are the indices corresponding to the upper and lower bound of the frequency band from 7 to 14 kHz.

LSD results with a 95% confidence interval for the three BWE methods are shown in Figure 5. Through Figure 5, we found that the LSD of the complicated audio signals is larger than the other types of audio signals. The reason is that many instruments are involved in complicated audio signals, and their energy in the HF spectrum is relatively high and not easy to accurately estimate. For simple audio signals, the energy attenuation of true HF components is obvious; thus, it may lead to low distortion. In addition, we can also find that the LSD of the PSR method is lower than EHBE and LE methods for complicated audio signals and singing audio signals. For simple audio signals, the LSD of the PSR method is a little higher. This may be because the energy attenuation of the extended HF spectrum in the PSR method is heavier than that of the true HF spectrum for harmonic-like signals.

An example of a symphony spectrogram reconstruction is shown in Figure 6. From the spectrogram, we can find that the spectra derived from LE and EHBE have some artifacts. This may cause the audio quality to be degraded. This result is identical to the objective evaluation.

Besides objective evaluation, our PSR method and the referential methods were applied into the G.722.1 WB audio codec [18] at 24 kb/s in order to further evaluate the subjective quality of BWE methods. The extended audio signals were compared with the SWB audio signals reproduced by G.722.1C [16] at the same bit rate by using a multiple stimuli with hidden reference and anchor (MUSHRA) listening test [19].

Fifteen male and five female expert listeners took part in the test with headphones. The test was arranged in a quiet room. The listening files in Additional file 1 include six audio signals, where five audio signals were used for test and one audio signal was used for training listeners. In MUSHRA listening test, the original audio signal was considered as a hidden reference and the low-pass filtered audio signal with bandwidth of 3.5 kHz was used as an anchor. The audio quality for G.722.1 codec, G.722.1C codec, LE method, EHBE method, and the proposed PSR method was evaluated for subjective listening test with a 100-point scale (100~80, excellent; 80~60, good; 60~40, fair; 40~20, poor; 20~0, bad). The

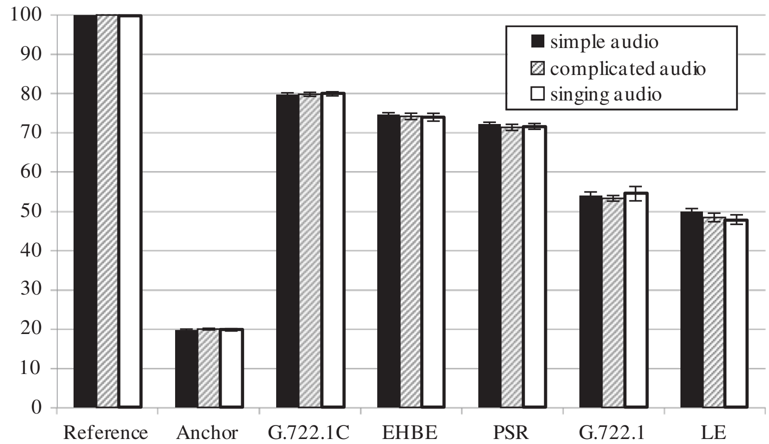


Figure 7 Mean subjective scores with 95% confidence intervals for the MUSHRA listening test.

result is illustrated in Figure 7 with a 95% confidence interval.

As shown in Figure 7, the subjective scores for the three different kinds of audio signals are similar in the MUSHRA test. In addition, the audio quality of G.722.1 with PSR is better than the original G.722.1 WB audio codec, but is inferior to the G.722.1C SWB audio codec. LE method shows its lower performance compared with PSR and EHBE methods. Both PSR and EHBE are preferred over the G.722.1 WB audio codec, while EHBE is better than the proposed PSR method on average. According to the informal listening tests, the tonal distortion which is caused by the non-linear filtering in EHBE is not sensitive to the listeners because the phase of the extended spectrum is comparatively continuous. Moreover, the PSR method can also reproduce the SWB audio signals whose subjective quality is similar to the EHBE method. So, the proposed PSR method is able to improve the quality of the WB audio signals decoded by G.722.1.

5 Conclusions

This paper presents a blind bandwidth extension (BWE) method of audio signals. Our BWE method consists of phase space reconstruction and high-frequency reconstruction of the audio signals. This blind BWE method incorporates both non-linear prediction and linear prediction. The objective tests show that the audio quality with the proposed method is a little bit better than the LE method and EHBE method. This BWE method was proved to be effective in quality enhancement of audio signals. In addition, the proposed method, LE method, and EHBE method were employed to extend the bandwidth of the audio signals decoded by the ITU-T G.722.1 WB codec. Subjective test results presented in this paper show that the proposed method greatly improved the auditory quality of the WB audio signals decoded by G.722.1. The bandwidth extension performance of the proposed method is higher than that of the typical LE method and is comparable to that of the conventional EHBE method.

Additional file

Additional file 1: This file contains audio samples.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under grant numbers 61072089 and 60872027, and the Beijing Natural Science Foundation Program and Scientific Research Key Program of Beijing Municipal Commission of Education (no. KZ201110005005). The authors wish to thank Dr. Christian Ritz, an associate professor in the University of Wollongong, for polishing the English language and for commenting on our paper. The authors also wish to thank the anonymous reviewers for their valuable comments that helped in significantly improving the presentation of our paper.

Received: 26 October 2012 Accepted: 28 November 2013
Published: 6 January 2014

References

1. E Larsen, RM Aarts, *Audio Bandwidth Extension: Application of Psychoacoustics, Signal Processing and Loudspeaker Design* (Wiley, New York, 2004)
2. M Dietz, L Liljeryd, K Kjörling, O Kunz, Spectral band replication, a novel approach in audio coding, in *Proceedings of the 112th AES Convention* (Munich, Germany, 2002)
3. P Ekstrand, Bandwidth extension of audio signals by spectral band replication, in *Proceedings of the 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio* (Leuven, Belgium, 2002)
4. International Standards Organization, *ISO/IEC 14496–3, Information technology – Coding of Audio-Visual Objects - Part 3: Audio* (International Standards Organization, Geneva, Switzerland, 2011)
5. S Meltzer, G Moser, MPEG-4 HE-AAC v2—audio coding for today's digital media world. EBU Technical Review **305**, 37–38 (2006)
6. 3GPP, *TS 26.290: Audio Codec Processing Functions—Extended Adaptive Multi-Rate-Wideband (AMR-WB+) codec; Transcoding functions* (3GPP, Sophia-Antipolis, 2004)
7. Z Jie, C Kihyun, O Eunmi, Bandwidth extension for China AVS-M standard, in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'09)* (Taipei, Taiwan, China, 2009)
8. CM Liu, WC Lee, HW Hsu, High frequency reconstruction for band-limited audio signals, in *Proceedings of the 6th International Conference on Digital Audio Effects (DAFX'03)* (London, UK, 2003)
9. C Budsabathon, A Nishihara, Bandwidth extension with hybrid signal extrapolation for audio coding. IEICE transactions on Fundamentals of Electronics, Communications and Sciences **90**(8), 1564–1569 (2007)
10. E Larsen, M Aarts, M Danessis, Efficient high-frequency bandwidth extension of music and speech, in *AES 112th Convention* (Munich, Germany, 2002)
11. Y-T Sha, C-C Bao, M-S Jia, X Liu, High frequency reconstruction of audio signal based on chaotic prediction theory, in *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP'10)* (Dallas, Texas, USA, 2010)
12. X Liu, C-C Bao, M-S Jia, Y-T Sha, Nonlinear bandwidth extension based on nearest-neighbor matching, in *Proceedings of the 2nd APSIPA ASC* (Biopolis, Singapore, 2010)
13. H Kantz, T Schreiber, *Nonlinear Time Series Analysis*, 2nd edn. (Cambridge University Press, Cambridge, 2004)
14. HDI Abarbanel, R Brown, JJ Sidorowich, LS Tsimring, The analysis of observed chaotic data in physical systems. Reviews of Modern Physic **65**(4), 1331–1392 (1993)
15. MH Hayes, *Statistical Digital Signal Processing and Modeling* (Wiley, New York, 1996)
16. International Telecommunication Union, *ITU-T Recommendation G.722.1 Annex C: Low Complexity Coding at 24 and 32 kb/s for Hands-Free Operation in Systems with Low Frame Loss Annex C 14 kHz Mode at 24, 32 and 48 kb/s* (International Telecommunication Union, Geneva, 2009a)
17. H Pulakka, L Laaksonen, M Vainio, J Pohjalainen, P Alku, Evaluation of an artificial speech bandwidth extension method in three languages. IEEE Trans. Audio Speech Lang. Processing **16**(6), 1124–1137 (2008)
18. International Telecommunication Union, *ITU-T Recommendation G.722.1, Low Complexity Coding at 24 and 32 kbit/s for Hands-Free Operation in Systems with Low Frame Loss* (International Telecommunication Union, Geneva, 2009b)
19. International Telecommunication Union, *ITU-T Recommendation BS.1543-1, Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)* (International Telecommunication Union, Geneva, 2001)

doi:10.1186/1687-4722-2014-1
Cite this article as: Bao *et al*.: A blind bandwidth extension method for audio signals based on phase space reconstruction. *EURASIP Journal on Audio, Speech, and Music Processing* 2014 **2014**:1.