Audio Bandwidth Extension Based on Grey Model

Haichuan Bai, Changchun Bao, Xin Liu, Hongrui Li
Speech and Audio Signal Processing Laboratory, School of Electronic Information and Control Engineering,
Beijing University of Technology, Beijing 100124, China
E-mail: baihaichuan@emails.bjut.edu.cn, baochch@bjut.edu.cn, liuxin0930@emails.bjut.edu.cn, lihongrui@emails.bjut.edu.cn

***Abstract*— A kind of audio bandwidth extension method based on Grey Model is proposed in this paper. Grey Model is utilized for estimating the envelope of high-frequency spectrum, according to the evolutionary tendency of audio spectral envelope. In addition, nearest-neighbor matching is utilized to predict the fine spectrum of high-frequency components. At last, through the envelope adjustment of high-frequency spectrum, the bandwidth extension of audio signals from wideband to super-wideband can be implemented. Objective performance evaluation indicates that the proposed method can effectively reconstruct the truncated high-frequency components and outperform the conventional method of audio bandwidth extension based on Gaussian mixture model.**

***Keywords*— Bandwidth extension, Grey Model, nearest-neighbor matching**

I. INTRODUCTION

Nowadays, wideband (WB) audio signals sampled at 16 kHz with a bandwidth of 7 kHz have a good intelligibility and timbre in practical applications. But the transparency and expressiveness of WB audio are not quite excellent due to the lack of the high-frequency (HF) components above 7 kHz. Therefore, super-wideband (SWB) audio signals sampled at 32 kHz with a bandwidth of 14 kHz is put forward in audio transmission. Bandwidth extension (BWE) can reconstruct the HF information from the WB audio signals at the decoder and enhance the auditory quality of the reproduced audio signals, without the modification of network transmission and source coding [1].

Traditional non-blind BWE methods usually transmit some side information describing the missing frequency bands and the timbre of reconstructed audio signals is pretty good. However, a great deal of side information will increase the burden of data transmission, that is, the coding efficiency of information source will be decreased. So the growing demands of modern communication cannot be satisfied by using the non-blind BWE methods. Therefore, blind BWE method is proposed to reconstruct the SWB audio signals without the additional information.

Blind BWE can be divided into two parts: the extension of spectral envelope and the extension of fine spectrum. Many typical methods of spectral envelope extension have been extensively studied, including linear extrapolation (LE) [2], codebook mapping [3], Gaussian mixture model (GMM) [4], Hidden Markov model [5] and so on. As a kind of classical statistic model, GMM has a good effect for smoothing or approximating some signals. By computing the joint probability density between WB audio features and HF spectral envelope energy, the HF sub-band energy could be estimated based on GMM in accordance with minimum mean square error (MMSE) criterion. However, due to the estimation error, the discontinuity may appear between the estimated HF spectral envelope and LF spectral envelope. It may lead to the degradation of the brightness and naturalness of the restored SWB audio signals.

In this paper, the BWE method based on Grey Model (GM) is introduced to predict the HF spectral envelope from the LF spectral envelope series, according to the exponential evolutionary tendency of audio spectral envelope. Meanwhile, the HF spectral fine structure is reconstructed based on nearest-neighbor matching (NNM) [6]. Eventually, BWE from WB to SWB audio signals can be implemented.

The paper is organized as follows: the estimation method of HF sub-band energy based on GM is described in Section 2. The overall framework of the proposed audio BWE method is illustrated in Section 3. In order to verify the effectiveness of algorithms, the objective quality test is presented in Section 4, and Section 5 draws the conclusions.

II. HIGH-FREQUENCY SUB-BAND ENERGY ESTIMATION BASED ON GREY MODEL

A. *Smoothness Analysis of the spectral envelope series of audio signals*

From the view of information theory and spectrogram analysis, there is a certain correlation and approximate trend between the HF and LF spectral envelope of the audio signals [2] [7]. In this paper, GM is introduced to establish a grey differential equation for describing the evolving trend of spectral envelope for both HF and LF components. In addition, the discarded HF spectrum components can be recovered by GM from the given WB spectral envelope energy and harmonic characteristics.

According to the grey system theory, the smoothness of the original series may directly affect the prediction accuracy of Grey Model [8]. Therefore, the smoothness analysis of the spectral envelope series of audio signals needs to be performed firstly.

The spectral envelope of the original WB audio signals can be represented by sub-band Root Mean Square (RMS) energy. The series $E^{(0)}$ that denotes the LF spectral envelope series of each frame of audio signals is given as follows:

$$E^{(0)}=\{e_{rms}{}^{(0)}(1), e_{rms}{}^{(0)}(2),\ldots, e_{rms}{}^{(0)}(N)\} \tag{1}$$

where $e_{rms}{}^{(0)}(k)\geq 0$, k=1,2,…,N, and N is the sub-band number of spectral envelope series of WB audio signals.

The smoothness ratio of the series $E^{(0)}$ is defined as:

$$\rho(k)=\frac{e_{rms}{}^{(0)}(k)}{\sum_{i=1}^{k-1} e_{rms}{}^{(0)}(i),} \qquad k=1,2,...,N \tag{2}$$

The spectral envelope series of WB audio signals $E^{(0)}$ is quasi-smooth series, if it satisfies:

a) $$\frac{\rho(k+1)}{\rho(k)}<1, \quad k=2,3,...,N-1 \tag{3}$$

b) $$\rho(k)\in[0,0.5), \quad k=3,4,...,N \tag{4}$$

Ten audio signals from MPEG database are selected in the experiment. These contain violin, drum, guitar, symphony and the length of each signal is 10~20s. The smoothness of spectral envelope series is analyzed afterwards. It is shown in Table I that most of the spectral envelope series satisfy the quasi-smooth conditions which are described in (3) and (4). Therefore, it can be considered that the spectral envelope series of audio signals have the characteristics of non-negativity and quasi-smoothness which are the basic requirements of GM [8].

TABLE I
SMOOTHNESS ANALYSIS OF SPECTRAL ENVELOPE SERIES

Audio	$\rho(k)\in[0,0.5)$ k=3,4,…,N	$\rho(k+1)/\rho(k)<1$ k=2,3,…,N-1
Symphony	98%	92%
Drum	95%	90%
Electronic organ	98%	93%
Guitar	91%	87%

B. *High-Frequency Sub-Band Energy Estimation*

According to the characteristics of non-negativity and quasi-smoothness, the GM with one order and single variable, namely GM(1,1), is used in this paper to predict spectral envelope. At first, an accumulated generating operation (AGO) is carried out on the original spectral envelope series, which can reduce the randomness of original series. Then, a grey differential equation can be set up by using the AGO series. At last, via the imitation of LF spectral envelope series, the prediction model is identified to predict the HF spectral envelope series. The principle of proposed method based on GM(1,1) is shown in Fig. 1:

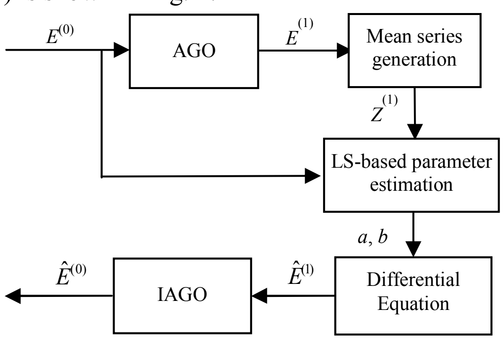


Fig. 1 Block Diagram of High-Frequency Sub-Band Energy Estimation Based on GM(1,1)

Firstly, the AGO series $E^{(1)}$ can be obtained according to the spectral envelope series $E^{(0)}$, and is written as

$$E^{(1)}=\{e_{rms}{}^{(1)}(1),\ e_{rms}{}^{(1)}(2),\ldots,\ e_{rms}{}^{(1)}(N)\} \tag{5}$$

where

$$e_{rms}{}^{(1)}(k)=\sum_{i=1}^{k} e_{rms}{}^{(0)}(i), \ k=1,2,...,N \tag{6}$$

The generating mean series $Z^{(1)}$ can be further derived from the above-mentioned series $E^{(1)}$ as,

$$Z^{(1)}=\{z^{(1)}(2),\ z^{(1)}(3),\ldots,\ z^{(1)}(N)\} \tag{7}$$

where $z^{(1)}(k)$ is the mean value of adjacent data, i.e.

$$z^{(1)}(k)=0.5e_{rms}{}^{(1)}(k)+0.5e_{rms}{}^{(1)}(k-1), \ k=2,3,...,N \tag{8}$$

Secondly, according to the series obtained, the grey differential equation describing the evolutionary tendency of spectral envelope series is defined as follows,

$$e_{rms}{}^{(0)}(k)+az^{(1)}(k)=b \tag{9}$$

where the parameters a and b are called the development coefficient and the grey input, respectively. The magnitude and sign of a reflect the evolutionary tendency of the predicted HF spectral envelope series $\hat{E}^{(1)}$ and the estimated original series $\hat{E}^{(0)}$. In addition, b denotes the evolutionary relationship between HF and LF spectral envelope series.

Thirdly, the parameters a and b in (9) can be calculated by the least square (LS) method.

Let $A=(a, b)^{T}$, provided that

$$Y=\begin{bmatrix} e_{rms}{}^{(0)}(2) \\ e_{rms}{}^{(0)}(3) \\ \vdots \\ e_{rms}{}^{(0)}(N) \end{bmatrix} \qquad B=\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(N) & 1 \end{bmatrix} \tag{10}$$

The estimated parameter A based on grey differential equation (8) is derived according to the LS method,

$$A=(B^{T}B)^{-1}B^{T}Y \tag{11}$$

In the grey system theory, the whitening equation of the grey differential equation (8) is defined as follows:

$$\frac{de_{rms}{}^{(1)}}{dt}+ae_{rms}{}^{(1)}=b \tag{12}$$

This equation is also called shadow equation and its time response function is expressed as:

$$e_{rms}{}^{(1)}(t)=(e_{rms}{}^{(1)}(1)-\frac{b}{a})e^{-at}+\frac{b}{a} \tag{13}$$

Therefore, the time response series of (9) can be obtained according to the related theorem [8].

$$e_{rms}{}^{(1)}(k+1)=(e_{rms}{}^{(1)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a}, k=N,N+1,...,N_{S} \tag{14}$$

where N_S+1 is the sub-band number of the spectral series of SWB audio signals.

Let initial value $e_{rms}{}^{(1)}(1)= e_{rms}{}^{(0)}(1)$ [8], the estimated value of AGO series $\hat{E}^{(1)}$ for HF spectral envelope is given by

$$\hat{e}_{rms}{}^{(1)}(k+1)=(e_{rms}{}^{(0)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a}, k=N,N+1,...,N_{S} \tag{15}$$

Finally, according to $\hat{E}^{(1)}$, the original series of HF spectral envelope of SWB audio signals is restored by the following equation:

$$\hat{e}_{rms}^{(0)}(k+1) = \hat{e}_{rms}^{(1)}(k+1) - \hat{e}_{rms}^{(1)}(k) = (1-e^{a})(e_{rms}^{(0)}(1) - \frac{b}{a})e^{-ak}, \; k = N, N+1, ..., N_S \tag{16}$$

By means of the procedures mentioned above, each value of HF sub-band energy can be predicted one by one. Combined with the original LF spectral envelope series, the proposed algorithm can effectively reconstruct the spectral envelope information of SWB audio signals.

III. AUDIO BANDWIDTH EXTENSION BASED ON GM(1,1)

The WB audio signal sampled at 16 kHz with an effective bandwidth of 7 kHz is used as the input signal in this paper. The principle of the proposed method is shown in Fig. 2:

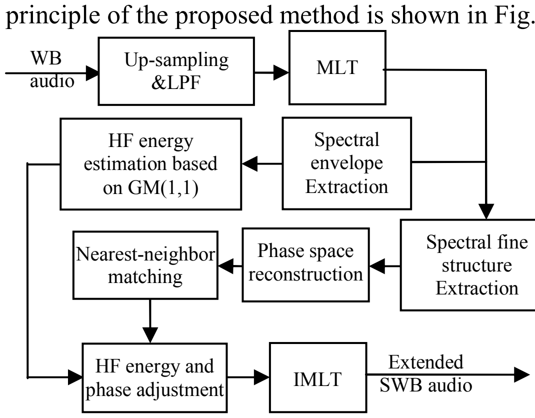


Fig. 2 Block Diagram of Proposed BWE Method

A. Time-frequency transform

The input audio signal is up-sampled and low-pass filtered, which could obtain a filtered signal re-sampled at 32 kHz. Then the Modulated Lapped Transform (MLT) with the overlap-adding length of 20ms is performed with the frame length of 20ms. The MLT coefficients above 7 kHz tend to be zero due to the low pass filter. Therefore, the corresponding MLT coefficients in the LF region are presented as $C_{mlt}(i)$, i=0…,279.

B. Bandwidth extension

Blind BWE method is composed of spectral envelope extension and spectral fine structure extension. Therefore, after the MLT on the input signal, these two parts will be processed in the proposed method, respectively.

a) Spectral envelope extension: The MLT coefficients below 7 kHz are divided into 14 sub-bands, and the root mean square (RMS) value of each sub-band can be written as: $e_{rms}^{(0)}(i)$, i=0,…,13. The non-negative series $E^{(0)}=\{e_{rms}^{(0)}(i), i=0,...,13\}$, is utilized to represent the LF spectral envelope information of each audio frame. GM(1,1) is used to predict the HF sub-band energy of audio signals according to $E^{(0)}$.

b) Spectral fine structure extension: The MLT coefficients below 7 kHz, $C_{mlt}(i)$, i=0…,279, are normalized by the root mean square value of each sub-band. The normalized MLT coefficients $C_{norm}(i)$, i=0,…,279 are used to represent the spectral fine structure information below 7kHz. According to the nonlinear characteristic of audio signals, $C_{norm}(i)$ can be used to reconstruct the phase space. HF phase points can be predicted from the trajectories of LF phase points by using nearest-neighbor matching (NNM) [6] to recover the HF spectral fine structure.

To begin with, the phase space is reconstructed and the m-dimensional phase points $S(i)$, i=0…279-(m-1)τ, can be presented as:

$$S(i) = \{C_{norm}(i), C_{norm}(i+\tau), C_{norm}(i+2\tau), ..., C_{norm}(i+(m-1)\tau)\} \tag{17}$$

where τ and m present the time delay and embedding dimension of audio spectral series in the current frame.

After reconstructing phase space, the trajectories of HF phase points can be predicted by NNM [6].

1) Compute the inner products between the new phase points $S(i)$, i=280-(m-1)τ and the LF phase point set $S=\{S(k)\}$, k=0…279-(m-1)τ.
2) The phase point $S(k_m)$ which maximizes the modulus of its inner product is chosen as the nearest neighbor of $S(i)$. Here, k_m is obtained by

$$k_m = \underset{k=0,...,279-(m-1)\tau}{\arg\max} \{|< S(i), S(k) >|\} \tag{18}$$

3) The predicted value of $C_{norm}(i+(m-1)\tau)$ is substituted by the highest-order element of $S(k_m)$.

The audio spectral series is updated through the above procedures, until the cutoff frequency of 14 kHz is reached up. At last, the HF components are recovered point by point.

C. Time-frequency inverse transform

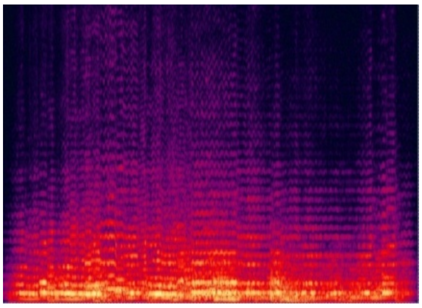
The HF spectral envelope restored by GM(1,1) is used for adjusting the energy of HF spectral fine structure and the truncated HF components is reconstructed eventually. In addition, combining with the original LF spectrum, the extended spectrum is transformed from frequency domain to time domain by an inverse MLT. Through the above steps, the overall BWE of the SWB audio signal is implemented.

IV. EVALUATION AND TEST RESULTS

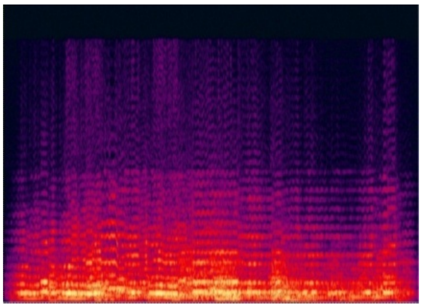
In order to evaluate the performance of proposed method, the spectrogram analysis and objective quality test are employed in comparison with the GMM-based method.

Audio signals from MPEG database are selected for tests, and all the audio signals are sampled at 32 kHz with the frame length of 20ms. These signals contain violin, drum, guitar, symphony, and so on. An example of the reconstructed spectrogram comparison about symphony signal is shown in Fig. 3.

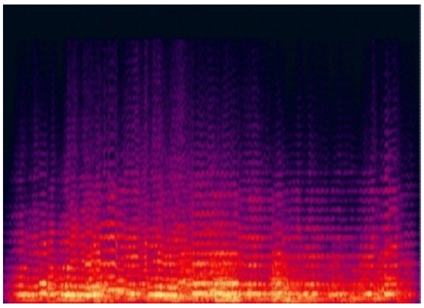
The spectrograms demonstrate that the proposed method can not only efficiently recover the HF spectrum, but also can keep the energy characteristics of audio spectrum. Compared with the original SWB audio, the HF energy of the audio extended by GMM is too smooth, and the transition part of HF and LF spectrum is connected unnaturally, so the auditory quality of reconstructed audio signals will inevitably be decreased. The proposed method overcomes these shortcomings. The extended spectrogram by GM(1,1) is much closer to the original one, and the naturalness of the restored SWB audio signals are enhanced obviously.



(a) Original SWB spectrogram



(b) Extended SWB spectrogram by GMM method



(c) Extended Spectrogram by the proposed method

Fig. 3 The reconstructed Spectrogram comparision about symphony signal

TABLE II
ODG VALUE COMPARISONS OF DIFFERENT METHODS

Audio samples	GMM	GM (1,1)
brahms_L.wav	-3.09	-2.755
dongwoo_L.wav	-2.655	-2.429
Music_1_s_L.wav	-2.798	-2.48
Music_2_L.wav	-2.962	-2.866
Music_3_L.wav	-3.291	-2.884
Music_4_s_L.wav	-3.523	-3.51
Music_5_s_L.wav	-3.39	-3.383
phi1_L.wav	-2.819	-2.803
phi2_L.wav	-3.545	-3.524
phi3_L.wav	-3.063	-2.968
Average scores	-3.114	-2.960

In objective audio quality measurement, perceptual evaluation of audio quality (PEAQ) [9] is performed to compare the proposed method with the GMM-based method. The scores of Objective Difference Grade (ODG) are given in Table II. It varies from -4 (very annoying) to 0 (imperceptible difference). The quality of reproduced audio signals achieves a significant improvement when ODG is increased by 0.1. In addition, the level of all the audio signals is equalized to -26dB and reference items are required to be up-sampled to 48 kHz.

Table II demonstrates that the average ODG score of audio signals reconstructed by GM(1,1) is -2.960, which has an improvement of 0.154 compared with GMM. This result indicates that the proposed algorithm is preferable over the conventional BWE method based on GMM.

V. CONCLUSIONS

In this paper, the BWE method from WB to SWB audio signals based on GM was proposed. GM(1,1) was utilized to predict the HF spectral envelope information from the LF spectral envelope series of audio signals. Moreover, the fine structure of the truncated HF spectrum is reconstructed by NNM. Finally, the high-quality SWB audio signals are reconstructed after adjusting the HF energy. The results of objective audio quality test indicate that the proposed method outperforms the traditional BWE method based on GMM.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 61072089, Grant No. 60872027).

REFERENCES

[1] Peter Vary, Rainer Martin, *Digital Speech Transmission-Enhancement, Coding and Error Concealme*nt. UK, John Wiley & Sons Ltd, 2006.

[2] Chig-Min Liu, Wen-Chieh Lee, Han-Wen Hsu, "High Frequency Reconstruction for Band-limited Audio Signals," *6th Conference on Digital Audio Effects*, London, UK, September, pp.1-6, 2003.

[3] Soon I.Y., Yeo C.K, "Bandwidth Extension of Narrowband Speech using Soft-decision Vector Quantization," *5th International Conference on Information Communications and Signal Processing*, pp. 734-738, 2005.

[4] Liu Xin, Bao Chang-chun, *et al*, "A Harmonic Bandwidth Extension based on Gaussian Mixture Model," *IEEE International Conference on Signal Processing*, pp. 474~477, 2010.

[5] Xin LIU, Chang-chun BAO, *et al*, "Nonlinear Bandwidth Extension of Audio Signals based on Hidden Markov Model," *IEEE International Symposium on Signal Processing and Information Technology*, Bilbao, Spain, December 14-17, pp. 144-149, 2011.

[6] X. Liu, C. Bao, M. Jia, and Y. Sha, "Nonlinear Bandwidth Extension based on Nearest-Neighbor Matching," *Processing of the 2nd APSIPA ASC*, Singapore, pp.169-172, 2010.

[7] Bernd Iser, Gerhard Schmidt, Wolfgang Minker. *Bandwidth Extension of Speech Signals*. USA, Springer-Verlag, 2008.

[8] Liu Si-feng, Lin Yi, *Grey Information-Theory and Practical Applications*. USA, Springer-Verlag, 2006.

[9] ITU-R Rec. BS.1387-1: Method for objective measurements of perceived audio quality. 1998-2001.