Audio Bandwidth Extension based on Maximum Lyapunov Prediction

Xingtao Zhang, Changchun Bao, Xin Liu, Liyan Zhang, Feng Bao and Bing Bu
Speech and Audio Signal Processing Laboratory, School of Electronic Information and Control Engineering,
Beijing University of Technology, Beijing 100124
E-mail: taozi@emails.bjut.edu.cn, baochch@bjut.edu.cn, liuxin0930@emails.bjut.edu.cn, zhangliyan1999@emails.bjut.edu.cn, baofeng0620@gmail.com, nono_bubing@emails.bjut.edu.cn

***Abstract*—In this paper, the maximum Lyapunov exponent and the correlation dimension are first utilized to determine the chaotic characteristics of audio spectrum based on nonlinear analysis method. Next, the maximum Lyapunov prediction and codebook mapping are used to recover the fine structure and envelope of high-frequency spectrum separately. By using above two approaches, the bandwidth of wideband audio is extended to super-wideband. Finally, this bandwidth extension method is applied into G.722.1 audio codec. The test results indicate that the proposed method outperforms the conventional algorithms in most cases, and the perceived quality of the extended super-wideband audio in the real wideband audio codec system is comparable with that of directly reconstructed audio using super-wideband audio coding at the bit-rate of 24 kb/s.**

I. INTRODUCTION

Due to the limitation of transmission bandwidth and storage capacity, usually the low-frequency (LF) components which is more perceptually important to human ear is transmitted in audio codec, and the less relevant high-frequency (HF) components is truncated during the transmission of audio signals. This inevitably leads to the degradation of audio quality, such as the timber depressing, the lack of loudness and naturalness. Obviously, the wideband (WB) audio can not satisfy people's need for high quality audio. And the transmission of high quality components at low bit-rate becomes the most urgent task in the field of audio coding. For this reason, audio bandwidth extension (BWE) emerges as the times require. There is a very little or no HF side information needed for transmission at the encoder with this method, and at the decoder, the truncated HF information can be efficiently reconstructed according to the relationship between the HF and LF information. Thus, the audio quality is improved to some extent.

According to whether HF side information is transmitted, there are mainly two BWE methods: the blind BWE [1][2] and the non-blind BWE. In the blind BWE method, no side information is transmitted at the encoder, and it can be employed in all audio codec. What's more, the quality of extended super-wideband (SWB) audio can be comparable with the quality of SWB audio signals. Therefore, it becomes the key part of audio codec. So far, the most efficient blind method is based on 'source-filter' model [3] which mainly includes two parts: the recovery of spectrum envelope and fine spectrum. Similar to BWE used in speech signal, it also has the same parts in audio BWE. But the recovery method of fine spectrum will not be suitable for audio signals because their characteristics are different.

In recent years, the research works on nonlinear theory are in rapid development, and have been applied to astronomy, hydrology, economy, and so on. It can predict the unknown part through the analysis of the known part of a series. It has been proved that the nonlinear prediction can keep high accuracy. For this reason, in this paper, the maximum Lyapunov prediction is used to reconstruct the HF components of audio signals, combined with codebook mapping to implement a blind audio BWE from WB to SWB.

The paper is organized as follows: The maximum Lyapunov exponent and the correlation dimension used for validating the nonlinear characteristics of audio spectrum series are discussed in section Ⅱ. The principles of audio BWE based on maximum Lyapunov prediction as well as the application to G.722.1 codec is described in Section Ⅲ. The objective quality test results are presented in Section Ⅳ and the conclusions are given in Section Ⅴ.

II. NONLINEAR CHARACTER ANALYSIS OF AUDIO SPECTRUM SERIES

The identification of chaotic characteristics is the prerequisite for chaotic study, and it mainly includes qualitative analysis, quantitative analysis, and the combination of the above two analyses [4]. The characteristics about stable motion, period motion, and chaotic motion are given in table I.

TABLE I
THE CHARACTERISTICS OF STABLE MOTION, PERIOD MOTION AND CHAOTIC MOTION

	Attractor	Correlation dimension D	Maximum Lyapunov exponent λ_1
Stable motion	Stable point	0	$\lambda_1 < 0$
Period motion	Limit cycle	1	$\lambda_1 \le 0$
Chaotic motion	Strange attractor	Non-integer	$\lambda_1 > 0$

At present, maximum Lyapunov exponent analysis and correlation dimension analysis are the major methods to verify the nonlinear characteristics of a series. As shown in Table I, when $\lambda_1 > 0$ and D is a non-integer, the chaotic characteristics of a series can be determined. The maximum Lyapunov exponent method has been used to verify the nonlinear characteristics in some literatures [5][6]. In this paper, the maximum Lyapunov exponent analysis is combined with the correlation dimension analysis to further verify the nonlinear characteristics of audio spectrum.

A. Maximum Lyapunov Exponent

There are many methods to get the maximum Lyapunov exponent, mainly including Bennetin method, Wolf method, Rosenstein small data method, Kantz small data method and so on [4]. Among them, the first two methods have the following defaults: a large computation; quite sensitive to the selected parameters. Due to the finite data, we adopt Rosenstein small data method to compute the maximum Lyapunov exponent in this paper.

The maximum Lyapunov exponent λ_1 can be estimated through the average divergence of adjacent trajectories in the basic trajectories. The definition of λ_1 is presented as

$$\lambda_1 = \frac{1}{t*N}\sum_{i=1}^{N}\ln\frac{d_i(t)}{d_i(0)} \tag{1}$$

where N is the number of adjacent trajectory pairs in phase space, $d_i(0)$ is the i^{th} Euclidean distance between two adjacent trajectories in the phase space at instant 0, and $d_i(t)$ is the i^{th} Euclidean distance between two adjacent trajectories in the phase space at instant t.

B. Correlation Dimension

The traditional method to calculate correlation dimension is G_P method [4]. By choosing arbitrary point $\mathbf{X}(i)$ of m-dimension phase space as reference phase point and calculating the distance between the reference phase point and other N-1 phase points, the number of points in volume element which takes $\mathbf{X}(i)$ as central point and small scalar r as radius is counted. Thus, the correlation integral (also called correlation function) $C_m(r)$ can be computed. The definition of $C_m(r)$ is given by:

$$C_m(r) = \frac{1}{N^2}\sum_{i\neq j} H(r - \|\mathbf{X}_i - \mathbf{X}_j\|) \tag{2}$$

where $H(\cdot)$ is Heaviside function.

Obviously, if radius r is too large or too small, the result will be influenced. When $r \to 0$, there exists the following relation:

$$\lim_{r\to 0} C_m(r) \propto r^D \tag{3}$$

where D is correlation dimension. By choosing a proper r which can make D describe the self-similar structure of strange attractor, the D becomes the slope of the $\ln C_m(r)$ versus $\ln r$ plot.

C. Nonlinear Characteristics Analysis of Audio Spectrum

TABLE II
STATISTIC RESULT OF λ_1 AND D OF DIFFERENT AUDIO SPECTRUM

Audio signals	Violin	Symphony	Drum	Guitar
Average λ_1	0.176	0.181	0.174	0.179
Average D	3.648	4.172	4.124	3.865

In this paper, the maximum Lyapunov exponent λ_1 and the correlation dimension D are calculated frame by frame. Audio signals for experiments are from Moving Picture Experts Group (MPEG) data set, including violin, symphony, drum and guitar. As depicted in Table II, all the λ_1 of the audio spectrum are positive and D is a non-integer. This result verifies the nonlinear characteristics of audio spectrum.

III. PREDICTION BASED ON MAXIMUM LYAPUNOV EXPONENT

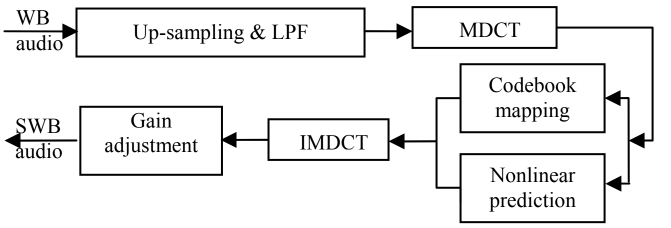

Fig. 1 The block diagram of the proposed method

The block diagram of the proposed method is shown in Fig.1. Firstly, WB audio signal sampled at 16 kHz is up-sampled by 2 and low-pass filtered. The filtered signal is re-sampled at 32 kHz with bandwidth of 7 kHz. This signal is transformed into frequency domain through Modified Discrete Cosine Transform (MDCT), and the LF MDCT coefficients are normalized by the root mean square value of each sub-band. Secondly, the HF MDCT coefficients are obtained based on maximum Lyapunov prediction and the HF envelope is adjusted with codebook mapping. Finally, the LF and HF components are combined in frequency domain, and transformed into time domain through inverse MDCT. With an appropriate gain adjustment, SWB audio signal is reproduced, and the blind BWE is finished. Among them, the nonlinear prediction module and codebook mapping module are the major modules which are described later in more detail.

A. Nonlinear Prediction

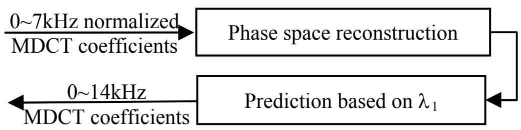

Fig. 2 The block diagram of nonlinear prediction

The block diagram of nonlinear prediction is depicted in Fig.2. Firstly, time delay method is used to reconstruct the phase space of stationary audio spectrum; and then, the maximum Lyapunov prediction method is adopted to get the HF spectrum.

- *Phase Space Reconstruction*

The main parameters used for phase space reconstruction include time delay τ and embedding dimension m. Only the proper reconstruction parameters are selected, the reconstructed phase space can be equivalent to the original nonlinear system, and then nonlinear prediction is made. Let $x(k)$, $k = 1, 2, \ldots, N$ as one-dimension spectrum series and N is a constant corresponding to the boundary index of LF and HF coefficients. The m-dimension phase vector $\mathbf{X}(n)$ obtained through phase space reconstruction is described as follows:

$$\mathbf{X}(n) = \{x(n), x(n+\tau), \ldots, x(n+(m-1)\tau)\},\quad n = 1,2,\ldots,N-(m-1)\tau \tag{4}$$

There are several methods to calculate the time delay τ, such as autocorrelation method, de-biasing autocorrelation method, average displacement, mutual information method and so on. Considering the simplicity and reliability of de-biasing autocorrelation method, it is selected to compute time delay τ.

The de-biasing autocorrelation function is presented as:

$$C_L(\tau)=\frac{1}{N-\tau}\sum_{n=1}^{N-\tau}(x_{n+\tau}-\bar{x})(x_n-\bar{x}) \tag{5}$$

where $\bar{x}$ is the mean of spectrum series. When the $C_L(\tau)$ is reduced to (1 – 1 / e) of $C_L(0)$, the corresponding τ is selected as the proper time delay of phase space.

There are many methods to calculate the embedding dimension m, such as False Nearest Neighbors (FNN), Cao method, G_P method and so on [4]. Where the selection of threshold of FNN method strongly depends on the subjective experience and the G_P method is fit for the phase space of intensively distributed points. Relatively, Cao method has the following advantages: only the parameter of delay time is needed, and the small data is enough for calculating the embedding dimension m. For this reason, Cao method is applied to compute embedding dimension m.

For the series x_1, x_2, …, x_N and phase points**,** the ratio of Euclidean distances between the phase point X(n) and its nearest neighbor point $\mathbf{X}^{NN}(n)$ is donated as $a(n,d)$ when the dimension number of the phase space is changed to d + 1 from d, i.e., $a(n,d)$ is defined by:

$$a(n,d)=\frac{\left\|\mathbf{X}_{d+1}(n)-\mathbf{X}_{d+1}^{NN}(n)\right\|}{\left\|\mathbf{X}_{d}(n)-\mathbf{X}_{d}^{NN}(n)\right\|},\quad n=1,2,...,N-d\tau \tag{6}$$

The mean of $a(n,d)$ is given by:

$$E(d)=\frac{1}{N-d\tau}\sum_{n=1}^{N-d\tau}a(n,d) \tag{7}$$

In order to get the minimum embedding dimension of the reconstructed phase space, $E_1(d)$ is defined as:

$$E_1(d)=E(d+1)/E(d) \tag{8}$$

If the variation of $E_1(d)$ stops at d_0, then d_0 is chosen as the proper embedding dimension m.

With the proper time delay τ and embedding dimension m, the phase space of audio spectrum series can be reconstructed, and the maximum Lyapunov exponent λ_1 is calculated.

- *Maximum Lyapunov Prediction*

Here, backward reconstruction method is introduced and m-dimension phase point $\mathbf{X}(n)$ can be represented as:

$$\mathbf{X}(n)=\{x(n),x(n-\tau),...,x(n-(m-1)\tau)\},\quad n=(m-1)\tau+1,(m-1)\tau+2,...,N \tag{9}$$

The phase point $\mathbf{X}(n)$ which makes $x(n+1)$ to be in the estimated phase point $\mathbf{X}(n+1)$ is chosen as prediction central phase point in the reconstructed phase space. Then the Euclidean distance $d_m(0)$ between $\mathbf{X}(n)$ and its nearest phase point $\mathbf{X}^{NN}(n)$ is given by:

$$d_m(0)=\min_{0\le k\le m}\left\|\mathbf{X}(n)-\mathbf{X}(n_k)\right\|=\left\|\mathbf{X}(n)-\mathbf{X}^{NN}(n)\right\| \tag{10}$$

According to the definition of λ_1, the distance between two phase points $\mathbf{X}(n)$ and $\mathbf{X}^{NN}(n)$ with one iteration will be:

$$\left\|\mathbf{X}(n+1)-\mathbf{X}^{NN}(n+1)\right\|=\left\|\mathbf{X}(n)-\mathbf{X}^{NN}(n)\right\|e^{\lambda_1} \tag{11}$$

In the above equation, only the last component of $\mathbf{X}(n+1)$ is unknown, therefore the value of $x(n+1)$ can be obtained.

In the proposed method, a group of phase points closing to the HF coefficients are chosen as central phase points. Firstly, m-dimension phase points are obtained through phase space reconstruction, and the nearest neighbor $\mathbf{X}(n_{NN})$ of the last phase point X(n), n = N - (m - 1)τ is searched at a proper interval. Then $\mathbf{X}(n_{NN})$ is used as prediction central phase point and the HF coefficients can be obtained based on maximum Lyapunov prediction given in equation (10). If $\mathbf{X}(n_{NN})$ reaches up to the boundary index of LF and HF coefficients , the iteration is stopped. Otherwise, let $n_{NN} = n_{NN} + 1$, and the nonlinear prediction is continued point by point until all the HF coefficients are obtained.

B. Adjustment of the High-Frequency Spectrum Envelope

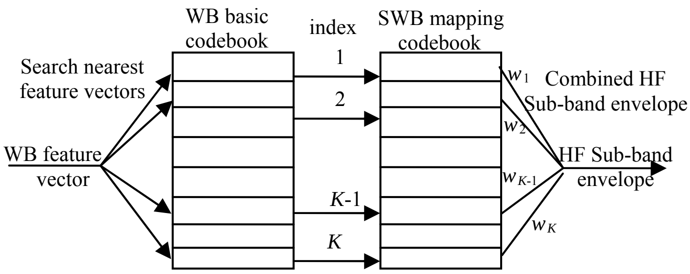

Fig. 3 The block diagram of codebook mapping

The block diagram of codebook mapping how to obtain the HF spectrum envelope from the LF spectrum is shown in Fig.3. Here, the WB features consist of sub-band envelope E_{rms}, sub-band flux F_{flux} and MPEG-7 timber features including audio spectrum centroid F_{ASC}, audio spectrum spread F_{ASS} and spectrum flatness F_{SF}. The fifteen-minute-long stationary audio signals are used for training the codebook based on LBG method.

The above WB features vector, $\mathbf{X} = \{F_{flux}, F_{ASC}, F_{ASS}, \mathbf{F}, \mathbf{E}\}^T$, in each audio frame is first computed, where $\mathbf{F} = \{F_{SF}(i), i = 7, 8, ..., 13\}^T$ and $\mathbf{E} = \{E_{rms}(i), i = 0, 1, ..., 13\}^T$. Then the HF spectrum envelope parameters are obtained based on fuzzy mapping method [7] with the following steps:

(1) Calculate the Euclidean distances between WB features vector $\mathbf{X}$ and WB basic code vectors, and search for K nearest code vectors in the distance set $\mathbf{D} = \{d_1, d_2, …, d_N\}$, then the distances are denoted as $\{d'_1, d'_2, …, d'_K\}$, where N = 256 and K = 8.

(2) Calculate the fuzzy membership of WB feature vector and its nearest code vectors by equation (12):

$$w_k=\left\{\sum_{l=1}^{k}\left(d'_k/d'_l\right)^{\frac{2}{m-1}}\right\}^{-1} \tag{12}$$

where m is the control parameter, here, m is set to 2. The fuzzy membership set is described as $\{w_1, w_2, …, w_K\}$.

(3) Estimate the HF sub-band envelope using SWB mapping codebook by equation (13):

$$\hat{\mathbf{E}}=\sum_{k=1}^{K}\frac{w_k}{\sum_{l=1}^{K}w_l}\mathbf{X}_k \tag{13}$$

Combining maximum Lyapunov prediction with codebook mapping, the truncated HF information during storage and transmission can be effectively recovered, and the audio quality will be improved.

C. Application of Bandwidth Extension in Audio Codec

Due to the limitation of transmission bandwidth, the audio quality obviously will degrade when only WB is transmitted. When the BWE method is used for audio enhancement module in decoder, the audio quality can be efficiently improved.

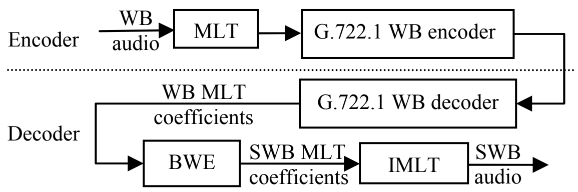

Fig. 4 The block diagram of BWE application in G.722.1 codec

The block diagram of BWE application in G.722.1 codec is shown in Fig.4. At the encoder, the input WB audio signal is transformed into frequency domain through Modulated Lapped Transform (MLT) and the MTL coefficients will be coded at proper bit-rate. At the decoder, the MLT coefficients firstly are decoded, and then SWB MLT coefficients can be obtained with BWE. Finally, SWB audio signals can be reconstructed by an inverse MLT.

IV. Evaluation and Test Results

In terms of objective audio quality measurement, the proposed algorithm is compared with two kinds of blind audio BWE methods, which are chaotic prediction [5] and Efficient High-frequency Bandwidth Extension (EHBE) [8]. In this paper, the testing audio signals are from MPEG data set and up-sampled to 48 kHz. The objective evaluation used in this paper is PEAQ test according to ITU-T BS.1387 [9]. The main parameter in PEAQ test is Objective Difference Grade (ODG) which varies from -4 (very annoying) to 0 (imperceptible difference). The change of ODG in the value of 0.1 is usually perceptually audible. Table Ⅲ shows the comparisons of ODG for the proposed method and other two blind methods.

Table Ⅲ
Comparisons of ODG Scores
of the Proposed Method and Other Two Blind Methods

	ODG			
	Violin	Symphony	Drum	Guitar
EHBE	-3.792	-3.725	-3.866	-3.664
Chaotic prediction	-2.993	-3.024	-3.174	-3.556
Proposed method	-2.804	-2.825	-2.878	-3.418

In addition, the extended SWB audio signals and directly reconstructed SWB audio signals are compared at the bit-rate of 24 kb/s. Table Ⅳ shows comparisons of ODG.

Table Ⅳ
Comparisons of ODG Scores
of the Reconstructed Audio between Two Methods

Audio signals	Directly reconstructed SWB audio signals in G.722.1c	Extended SWB audio signals in G.722.1
Violin	-3.646	-3.656
Symphony	-3.387	-3.345
Drum	-3.743	-3.621
Guitar	-3.742	-3.683

The test results indicate that the proposed method outperforms the other two blind BWE algorithms, and the perceived quality of the extended SWB audio in the real WB audio codec system is comparable with directly reconstructed SWB audio at the same bit-rate in G.722.1c audio codec.

V. Conclusions

In this paper, the chaotic characteristics of audio spectrum are determined by the maximum Lyapunov exponent and the correlation dimension. The maximum Lyapunov prediction method is adopted to recover the high-frequency spectrum, and combined with codebook mapping to adjust the high-frequency spectrum envelope, thus the blind bandwidth extension from wideband to super-wideband is finished. Furthermore, the bandwidth extension is applied into G.722.1 codec. Test results indicate that the proposed method outperforms the conventional bandwidth extension algorithms, and the perceived quality of the extended super-wideband audio signals in G.722.1 wideband audio codec is comparable with directly reconstructed SWB audio signals at the bit-rate of 24 kb/s in G.722.1c super-wideband coder in most cases.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 61072089, Grant No. 60872027), the Beijing Municipal Natural Science Foundation (Grant No. 4082006), the Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality, the Doctoral Research Initiation Foundation Project of Beijing University of Technology and the 9th Postgraduate Science Foundation of Beijing University of Technology (ykj-2011-4893).

References

[1] Y. T. Sha, C. C. Bao and M. S. Jia, "A method of high frequencies reconstruction of audio signal based on reconstructed octave," *Signal Processing*, vol. 25, 2009, pp. 139-142. (in Chinese).

[2] X. Liu, C. C. Bao, M. S. Jia and Y. T. Sha, "A harmonic bandwidth extension based on Gaussian Mixture Model," *ICSP2010*, Beijing, CHINA, October, 2010, pp. 474-477.

[3] C. C. Bao, *Principles of Digital Speech Coding*. Xi'an: Xidian University Press, 2007. (in Chinese)

[4] M. Han, *Prediction Theory and Method of Chaotic Time Series*. Beijing: China WaterPower Press, 2007. (in Chinese).

[5] Y. T. Sha, C. C. Bao, M. S. Jia and X. Liu, "High frequency reconstruction of audio signal based on chaotic prediction theory," *10^{th} International Conference on Signal Processing (ICSP2010)*, Dallas, Texas, USA, March 14-19, 2010, pp. 381-384.

[6] X. Liu, C. C. Bao, M. S. Jia and Y. T. Sha, "Nonlinear bandwidth extension based on nearest-neighbor matching," *Processing of the Second APSIPA ASC*, 2010, pp. 169-172.

[7] Y. Zhang and R. M. Hu, "Speech wideband extension based on Gaussian Mixture Model," *Acta acustica*, September, 2009, 34(5), pp. 471-480. (in Chinese).

[8] C. M. Liu, W. C. Lee and H. W. Hsu, "High frequency reconstruction for band-limited audio signals," *Proc. of the 6th Int. Conference on Digital Audio Effects. (DAFX)*, London, UK, September 8-11, 2003.

[9] ITU-R Rec. BS.1387-1, *Method for Objective Measurements of Perceived Audio Quality*, 1998-2001.