HIGH FREQUENCY RECONSTRUCTION OF AUDIO SIGNAL BASED ON CHAOTIC PREDICTION THEORY

Yong-tao SHA   Chang-chun BAO   Mao-shen JIA   Xin Liu

Speech and Audio Signal Processing Laboratory, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China
shayt@emails.bjut.edu.cn, baochch@bjut.edu.cn, jiamaoshen@emails.bjut.edu.cn, liuxin0930@emails.bjut.edu.cn

ABSTRACT

The quality of audio signals that have been encoded with low-bit rate audio coding standards is degraded because the high frequency information has been removed. The quality of such audio signals can, however, be improved by reconstructing the high frequency information which was lost. In this paper the principles of audio signal production and the characteristics of the human hearing system have been used to develop a blind high frequency reconstruction method based on chaotic prediction theory. Performance evaluation with objective and subjective tests has shown that this method is, in most cases, more efficient than other blind high frequency reconstruction methods.

Index Terms— Audio coding, Non-linear systems, High Frequency Reconstruction, Chaotic prediction

1. INTRODUCTION

Low bit-rate audio coding methods remove the less perceptually relevant high frequency information in order to reduce the amount of data needed to represent the audio signal. The quality of the resulting signal is somewhat diminished due to a loss of brightness, naturalness and timbre.

The quality of the signal can be improved by using a high frequency reconstruction technique which is able to recover some of the lost information. This method has become a key to improving the coding efficiency whilst keeping the bit rate requirements low. High frequency reconstruction is now widely employed in many audio coding standards - for example, MPEG-4 High Efficiency Advanced Audio Coding (HE-AAC) is a combination of MPEG-4 AAC and Spectral Band Replication (SBR) [1]. For a given bit rate the quality of a signal encoded using HE-AAC is higher than the equivalent AAC encoded signal.

SBR is what is known as a non-blind high frequency reconstruction technique which contains side information related to high frequencies. On the other hand, blind high frequency reconstruction technique does not use such side information, reducing the number of bits required; they are also not tied to a particular codec. Blind methods employ a variety of ways to obtain the high frequency information: for example, it can be reconstructed by direct replication from the low frequency information [2], by using a non-linear device [3], or by linear prediction [4].

This paper presents a blind high frequency reconstruction method based on chaotic prediction theory. In this method, the high frequency information is reconstructed according to the principles of audio production and the characteristics of the human ear.

The paper is organised in the following way: section 2 discusses the chaotic characteristics of audio signals; section 3 describes how chaotic prediction theory can be used to reconstruct high frequency information; section 4 evaluates the performance of the proposed algorithm and the conclusion is given in section 5.

2. ANALYSIS OF THE CHAOTIC CHARACTERISTICS OF AUDIO SIGNALS

An audio source which is not derived from human speech is usually composed of the output from several instruments. The audio signal varies from moment to moment and the associated spectrum follows different change laws - fundamentally the audio signal is non-linear in both the time domain and the frequency domain.

Time series is the conventional method for modeling and processing non-linear systems, but audio-signals are traditionally manipulated in the frequency domain. This paper presents a new idea: if the whole frequency spectrum of the audio signal is seen as a section of a time series then non-linear prediction can be used to obtain the high frequency information. The non-linear dynamics theory used for time series is also an efficient tool for processing frequency series.

It has been proven in [5] that speech contains chaotic characteristics. In this paper the Lyapunov exponent [6] is used to prove that the audio spectrum contains chaotic characteristics – it can describe the growth rate or shrinkage rate of small perturbations in different directions in phase space. The audio spectrum can be characterised as chaotic if any one of the Lyapunov exponents is positive - only the maximum Lyapunov exponent needs to be considered. This paper uses an efficient method [7], which is relatively easy to calculate and requires only a small amount of data. The maximum Lyapunov exponent, λ_1, is calculated from equation (1)

$$\lambda_1 = \frac{1}{t \cdot N} \sum_{j=1}^{N} \ln \frac{d_j(t)}{d_j(0)}, \tag{1}$$

where N is the total number of data samples, $d_j(0)$ is the j^{th} initial distance between two adjacent vectors in the phase space, and $d_j(t)$ is the distance at instant t.

Audio output signals derived from a violin, a drum, a guitar, an harmonica and a symphony orchestra were used to prove that the audio spectrum has chaotic characteristics. For each frame the maximum Lyapunov exponent λ_1 of the frequency coefficients is calculated. Table 1 shows the percentage, φ, of positive λ_1 and the average, δ, of λ_1 for each audio signal.

Table 1 Statistical results for λ_1

Audio / Parameter	Violin	Drum	Guitar	Harmonica	Symphony
φ	100%	100%	97.05%	99.24%	100%
δ	0.122	0.125	0.120	0.123	0.126

As can be seen from Table 1 almost all the λ_1 are positive i.e. the audio spectrum really does have chaotic characteristics.

3. HIGH FREQUENCY RECONSTRUCTION METHOD

A Modified Discrete Cosine Transform (MDCT) is used to calculate the frequency coefficients. This transform is more suitable to be used to reconstruct high frequency information from an audio signal than the Fast Fourier Transform (FFT) or the Discrete Cosine Transform (DCT). Unlike the FFT, it does not contain any phase information, and compared to both the FFT and DCT it can reduce the blocking effect. The proposed method is shown in Figure 1.

3.1. Phase space reconstruction of the audio signal's spectrum information

The phase space of the MCDT coefficients' series should be reconstructed using a non-linear system because they have chaotic characteristics. The information hidden in the MDCT coefficients' series can be recovered when the series is expanded from one-dimensional space to multi-dimensional space. Phase space reconstruction is a prerequisite for high frequency reconstruction.

F. Takens and N. H. Packard [8] have proposed a method to reconstruct the phase space from one-dimensional space. Based upon this method, only the low frequency MDCT coefficients are used to reconstruct the phase space as the high frequency MDCT coefficients have been lost.

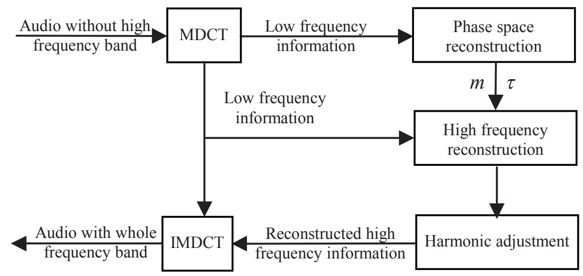

Figure 1. Block diagram of the proposed method

The low frequency MDCT coefficients' series is defined as $\{x(k),\ k = 1, 2, \dots, M\}$, where k is the index of MDCT coefficients and M is a constant corresponding to the boundary index of low and high frequency MDCT coefficients. The embedded dimension, m, and the embedded delay, τ, can be calculated using conventional methods [9]. The reconstructed feature vector is $\mathbf{y}(k)=\{x(k), x(k+\tau), x(k+2\tau),\dots, x[k+(m-1)\tau]\}$, where $k = 1, 2, \dots, M - (m - 1)$. Thus, the dynamic laws which describe the whole system are represented by the distribution of the trajectory in phase space.

It may be that each frame comes from a different system, so the dynamical laws which the MDCT coefficients follow may differ – thus the phase space characteristics will also be different. In order to reconstruct the frame's phase space both m and τ need to be calculated for each frame.

3.2. Non-linear prediction of the high frequency MDCT coefficients

The lost high frequency MDCT coefficients are recovered by using non-linear prediction, after the low frequency MDCT coefficients' phase space has been reconstructed.

A local prediction method, based on an adaptive algorithm, is used to predict the high frequency MDCT coefficients. Local prediction has two advantages over global prediction: a smaller number of calculations need to be made, and it is much more accurate because the calculation is based on neighbouring points which have similar characteristics. Takens' Theory [8] has shown the equivalence between the distribution of the trajectory in phase space and the original dynamics of the system. There is a non-linear mapping between the phase points in phase space – it is this that is used to predict the high frequency information. The specific method is described as follows:

(1) A group of low frequency MDCT feature vectors which are closest to the boundary index M is chosen as the predictive central vectors. This avoids the inaccuracies that would result in if a single predictive central vector was chosen. Region 0 in Figure 2 below, shows the highest dimensional components of all the predictive central vectors. The number of predictive central vectors is usually set to 2τ.

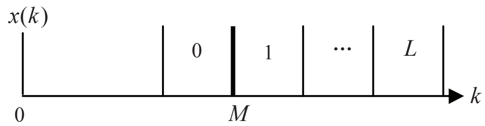

Figure 2. A sample of predictive central vector selection

(2) The Euclidean distance is used to find the neighbours of each predictive central vector. The number of neighbours is chosen to be in the interval [$2m$, $2m + 10$] – choosing too many neighbours or too few neighbours is detrimental to the calculation.

(3) The high frequency band is divided into L regions by 2τ, where region L contains the highest frequency shown in Fig. 2. The MDCT coefficients in regions 1 to L should be predicted by the corresponding predictive central vectors. The non-linear predictive model based on an adaptive algorithm is given by

$$x(k_c + 2\tau \cdot l) = \sum_{i=0}^{m-1} w_0(i) h\left[x(k_c - i\tau)\right] + \sum_{i=0}^{m-1} w_1(i) x(k_c - i\tau) + C\,, \quad (2)$$

where $k_c = M - 2\tau + 1, M - 2\tau + 2, \ldots, M$, represents the position of the highest dimensional component of the predictive central vector; $l = 1, 2, \ldots, L$, is the region index of high frequency band; C is a constant; $w_0(i)$ is the non-linear weight coefficient; $w_1(i)$ is the linear weight coefficient, and $h(i)$ is a non-linear kernel function. The actual function used is

$$h(i) = \frac{1}{1 + e^{-i}}\,. \quad (3)$$

(4) The neighbours of the current predictive central vector which lies at a distance of $2\tau \cdot l$ from some high frequency MDCT coefficient to be predicted should be chosen as predictive central vectors. The predicted MDCT coefficients which lie at a distance of $2\tau \cdot l$ from the highest dimensional component of these neighbours are determined. By considering these neighbours and the predicted MDCT coefficients into equation (2), the weight vector $W^T(n)$ are calculated using the weighted Recursive Least Squares (RLS) by obtaining the minimal cost function from equation (4). $W^T(n)$ reflects the relationship between the predictive central vector and the predicted high frequency MDCT coefficients.

$$\varepsilon(n) = \sum_{i=1}^{n} \gamma^{n-i} \left\{W_r\left[d(i) - W^T(n) Y(i)\right]\right\}^2\,, \quad (4)$$

where n is the number of neighbours; γ is the 'forgetting factor'; $d(i)$ is the actual value of the i^{th} determined MDCT coefficient;

$W^T(n) = \{\ w_0(0)\ ,\ldots,\ w_0(m-1)\ ,\ w_1(0)\ ,\ldots,\ w_1(m-1)\ ,\ C\ \}$;

$Y(i) = \{\ h[x(i)], \ldots, h[x(i-(m-1)\tau)]\,,\ x(i)\,, \ldots, x(i-(m-1)\tau)\ ,\ 1\ \}$,

and $W_r = e^{-(R_r - R_{\min})} \Big/ \sum_{i=1}^{n} e^{-(R_i - R_{\min})}$, where R_r is the Euclidean distance between the r^{th} neighbour and the predictive central vector, $R_{\min}$ is the minimal Euclidean distance.

The accuracy of the prediction can be improved by introducing W_r, according to the following rule: the nearest neighbours which have similar characteristics to the central vector should have a large effect on the result of the prediction.

(5) After the weight vector $W^T(n)$ has been obtained, equation (2) is used to calculate the high frequency MDCT coefficients located at a distance of $2\tau \cdot l$ from the highest dimensional component of the predictive central vector.

In practice the predictive central vector is adjusted by changing the value of k_c, starting from a value of $l = 1$. Equation (2) is used to calculate each high frequency MDCT coefficient which lies at a distance of $2\tau \cdot l$ from the highest dimensional component of the corresponding predictive central vector. All the high frequency MDCT coefficients in region 1 (see Fig. 2) can be calculated by changing the value of k_c. Thereafter, the MDCT coefficients in regions 2 to L are obtained by using a step size of 1 for l.

(6) The high frequency MDCT coefficients need to be adjusted using the sub-band energy in order to attenuate the noise and to reduce the effects which are cause by large prediction errors.

The energy adjustment is set to be equal to the envelope adjustment. The total frequency band is divided into N_{sub} sub-bands. The envelope $B_{rms}(i)$ of each sub-band is calculated by

$$B_{rms}(i) = \sqrt{\frac{1}{M_{sub}} \sum_{j=0}^{M_{sub}-1} x^2(M_{sub} \cdot i + j)}\,, \quad (5)$$

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

The ratio of envelope between two adjacent sub-bands $\eta(i)$ can be obtained by

$$\eta(i) = \frac{B_{rms}(i)}{B_{rms}(i-1)}\,, i = 1, 2, \cdots N_{sub} - 1\,. \quad (6)$$

The average, A_{avg} , of all $\eta(i)$ between two adjacent low frequency sub-bands is calculated. The noise can be reduced by diminishing the envelope. If $A_{avg} > D$ (e.g. $D = 0.95$), then let $A_{avg} = D$.

In practice, once the high frequency MDCT coefficients in a sub-band have been obtained, they need to be adjusted from one sub-band to another using equation (7) below.

$$\hat{x}(M_{sub} \cdot i + j) = x(M_{sub} \cdot i + j) \cdot \frac{A_{avg}}{\eta(i)} \quad (7)$$

3.3. Adjusting the high frequency harmonic components

It has been proven by experiment that audio quality can be improved by introducing harmonic components into the high frequency spectrum. The harmonic components in the high frequency spectrum need to be adjusted because the harmonic components are not, by themselves, sufficient to improve the audio quality. The process is described below.

(1) The FFT is used to transform the high frequency and low frequency MDCT coefficients, respectively;

(2) The amplitude information from some of the larger FFT coefficients of low frequency MDCT coefficients (1-2 % of the total) is used to replace the FFT coefficients of high frequency MDCT coefficients at the same frequency index;

(3) The Inverse FFT is then used to transform the high frequency FFT coefficients back to the MDCT domain;

(4) The last step is to adjust the high frequency energy using Step (6) in section 3.2.

4. PERFORMANCE EVALUATION

The test audio samples used in this paper were provided by the MPEG standards' organization. The audio sampling frequency of the test files was 28 kHz: the low frequency band was 0~7 kHz whilst the high frequency band was 7~14 kHz. The frame length was 20 ms. The audio signals were from a number of different instruments: a violin, a drum, a guitar, an harmonica and a symphony orchestra. The proposed method was evaluated using both objective and subjective tests.

The objective evaluation test used in this paper was the PEAQ test from ITU-T BS.1387 [10]. The main parameter in the PEAQ test is the Objective Difference Grade (ODG). Audio signals should be up-sampled to 48 kHz in order to compute the ODG. The ODG parameter can take on a range of values, from -4 (very annoying) to 0 (imperceptible impairment). A difference of 0.1 in the ODG value is audibly perceptible.

Table 2 shows the value of the ODG parameter for the proposed method and the methods described in [2] and [3].

Table 2 Values of the ODG parameter for the three algorithms

Method	ODG value				
	Violin	Drum	Guitar	Harmonica	Symphony
Proposed method	-3.768	-3.570	-3.647	-3.877	-3.012
Method in [2]	-3.856	-3.627	-3.779	-3.878	-3.739
Method in [3]	-3.862	-3.828	-3.731	-3.786	-3.788

It can be seen from Table 2 that for 4 out of the 5 audio samples, the method described in this paper is more accurate than the methods described in [2] and [3].

Table 3 Average MOS values for the three algorithms

Method	Average MOS value				
	Violin	Drum	Guitar	Harmonica	Symphony
Proposed method	3.83	3.67	3.75	3.25	3.58
Method in [2]	2.67	2.17	2.58	3.17	2.08
Method in [3]	3.33	3.17	3.17	3.33	2.83

The subjective evaluation test used in this paper was the Mean Opinion Score (MOS) test. Fifteen people were invited to take a test in which the decoded audio for the proposed method and the methods described in [2] and [3] were compared, with the original audio as a reference. Table 3 shows the average MOS value for the three algorithms.

It can be seen from Table 3 that the method described in this paper produces the best MOS score for 4 out of the 5 audio samples.

5. CONCLUSION

In this paper the principles of audio signal production and the characteristics of the human hearing system have been used to develop a blind high frequency reconstruction method based on chaotic prediction theory. The proposed method is independent of the audio codec. Performance evaluation with objective and subjective tests has shown that this method is, in most cases, more efficient than other blind high frequency reconstruction methods.

6. ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (grant no. 60872027), the Beijing Natural Science Foundation (grant no. 4082006) and Beijing University of Technology's Innovation Project for Ph. D Candidates (bcx-2009-014).

7. REFERENCES

[1] Martin Wolters, et al. A closer look into MPEG-4 High Efficiency AAC. 115th AES Convention, New York, USA, October 10–13, 2003, 1-16.

[2] Chi-Min Liu, et al. High frequency reconstruction for band-limited audio signals. Proc. of the 6th Int. Conference on Digital Audio Effects, London, UK, September 8-11, 2003, 1-6.

[3] Erik Larsen, et al. Efficient high-frequency bandwidth extension of music and speech. 112th AES Convention, Munich, Germany, May 10-13, 2002, 1-5.

[4] Chatree Budsabathon, et al. Bandwidth Extension with Hybrid Signal Extrapolation for Audio Coding. IEICE TRANS. Fundamentals, August, 2007, Vol.E90-A, No.8, 1564-1569.

[5] QIN Ai- na, et al. Nonlinear speech predictor using models for chaotic systems. Computer Engineering and Applications, 2008, 44(18), 141- 143. (in Chinese)

[6] Rasband S N. Chaotic Dynamics of Nonlinear Systems. New York: John Wiley & Sons, 1990.

[7] Rosenstein M T, et al. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D, 1993, 65(1-2), 117-134.

[8] Takens F. Detecting strange attractors in turbulence. In Lecture Notes in Math, Vol 898. Berlin: Springer, 1981. 366-381.

[9] CHEN Keng, HAN Bo-Tang. A Survey of State Space Reconstruction of Chaotic Time Series Analysis. COMPUTER SCIENCE, 2005, 32(4), 67-70. (in Chinese)

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