**Applied Mathematics & Information Sciences**
*An International Journal*

# Research of Spatial Parameter Codec in MDCT Domain

*Cao Sheng*[1],[3] *, Hu Ruimin*[1] *, Peng Yuxing*[2] *and Liu Yutian*[1]

[1]National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan 430079, China
[2]National University of Defense Technology, Changsha 410073, China
[3]Hubei Telecom Billing Center, Wuhan 430022, China

**Abstract:** With stereo audio stepping into people's daily lives, spatial parametric technology comes to its broad prospects for development. Spatial parameteric technology is a method for stereo compression. MDCT is excluded in the current scheme for spatial cues representation, due to its lacking of phase information and energy conservation. In this paper, Construction of new spatial parameter codec framework in MDCT domain, and Optimization of calculating for the new module, to reduce computational complexity of the system.Through the research, implementation, and testing of this two parts, the new spatial parameter codec system, which is constructed in this subject, completes all the basic functions of spatial parametric technology. To compare with the EaacPlus system, the output stereo of the new system gains a considerable subjective and objective test scores. This research makes a contribution to the spatial parametric technology in MDCT domain, and provides a new train of thought to reduce complexity of EaacPlus system.

**Keywords:** MDCT, Complesity, Stereo, Compression, Spatial parameters

## 1 Introduction

The spatial parameter is a key point of the spatial audio coding technology. The core of spatial audio coding is to compress multi-channel audio into a mono signal to achieve the purpose of significantly reducing the transmission rate.

The spatial parameters attached to the space psychoacoustic theory of the formation. The Spatial Psychoacoustic [1] is a study for the environmental characteristics of the human ear perception of sound field. And the human ear perception is not obvious for the other spatial information, which to be used or not, taking into account the coding efficiency of this article codec to discard such low utility of spatial information.

There are three spatial parameters [2,3] to describe the spatial information that is selected as the control parameter digital flow size for this codec. Interaural Level Difference (ILD), the size ratio of said left and right channels of sound pressure level is used to distinguish the horizontal direction of the sound source angle. Interaural Time Difference (ITD) between the ears, it is another can represent the spatial parameters of the sound source azimuth, said the delay size of the left and right channels. Interaural Coherence (IC) between the ears, it is an important parameter to increase a signal sense of space. At present, the mature techniques of the space parameters coding are used by time domain[4]. Extracting spatial parameters of time domain integrity to retain their physical meaning, and thus is the entry point of the spatial parameters has the advantages of accurate and reliable.

Further improved the accuracy of the spatial parameters encoded spatial parameters are divided into independently extracted at different frequencies[5], with full consideration to the role of different size in different frequency band-sial parameters. This approach comes from studies of the human ear, the signal-processing point of view. The physiological structure of the human ear is a natural cochlea filter, and the human ear is divided into the number of frequency sub-band, relatively independent of each sub-band acoustic reception and perception[6]. Therefore, do frequency division and then extracting spatial parameters, in line with the basic physical structure of the human ear, and thus a more accurate description of the spatial orientation information to enhance the quality of the encoder.

In summary, the sub-band coding method in time domain for the space parameter is more mature and sound quality good and the spatial audio coding strategy. This

---

* Corresponding author e-mail: caosheng@whu.edu.cn

article as a basis for further research is to seek the expansion and creation of space parameters.

## 2 SPATIAL PARAMETERS IN MDCT DOMAIN

Spatial parameters from time domain to the MDCT domain for the following three considerations:

a) The spatial parameter encoding and sophisticated audio technology combined. Digital signal processing technology, to make a variety of time-frequency transform technology into actual coding. Such as Popular MP3, AAC encoded by virtue of MDCT approach to obtain good results. In order to broaden the use of spatial parameters, the study of the technical points necessary to move closer to the forefront of the audio coding technology. MDCT is a good platform for the new technology offers vast opportunities for development.

b) Replace the QMF filter to reduce complexity, increase flexibility of Frequency Allocations. Evolution of the cochlear filter QMF filter technically has made considerable progress. However, the QMF filter has to face the division flexibility and complex through the high fatal flaw. MDCT transform exactly in this area can restore a disadvantage. Compare their performance, QMF filter for every different frequency division needs to re-design, while MDCT directly to the signal with time-frequency transformed. In computational complexity, MDCT is a mature transformation method, has developed a fast algorithm. The coding system of computing speed will be improved dramatically after replacing the QMF MDCT.

c) Space parameters of the MDCT-domain space and time-domain parameters associated for theory development. Between the spatial parameters of the MDCT-domain and time-domain spatial parameters, the space parameter is very similar. For example, ILD parameter three dimensional distribution in time-domain and MDCT domain shown in Figure 1.
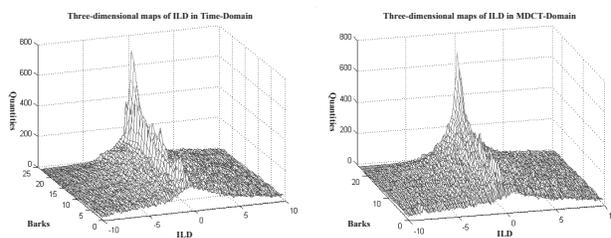


**Fig. 1** Three-dimensional maps of ILD parameter in time-domain and MDCT-domain

In Figure 1, left the ILD parameter is in the time-domain three-dimensional statistical distribution and the right by the ILD parameter in MDCT-domain three-dimensional statistical distribution. Three directions, horizontal level indicates that the parameter the ILD numerical size; the level of vertical, said the serial number of the frequency allocation band, here the signal is divided into 24 frequency sub-band; the vertical direction by a large number of audio data statistics, the statistics under the respective conditions number. Parameters by the ILD-frequency three-dimensional maps of the two domains can be seen, both are very similar to numerical size from the distribution shape, proving once again that when the frequency of the two spatial parameters of the domain interoperability. Similarly, the ITD, the IC parameter also has a similar nature.

## 3 SPATIAL PARAMETER CODEC SYSTEM IN MDCT DOMAIN

### 3.1 SPCS ARCHITECTURE IN MDCT DOMAIN

Parametric Stereo QMF as a means of frequency division is nowadays the mainstream space parameter encoding method [7,8,9,10]. The method is an excellent representative of in the time domain to extract spatial parameters, and has been adopted by the MPEG standard [11,12,13,14]. The parameters of MDCT-domain coding is intended as a basis, to make space parameter-related modifications in order to achieve a new type of spatial parameters of the codec. The original time-domain spatial parameters encoded block diagram shown in Figure 2. It modified the coding of spatial parameters of the MDCT-domain block diagram shown in Figure 3.
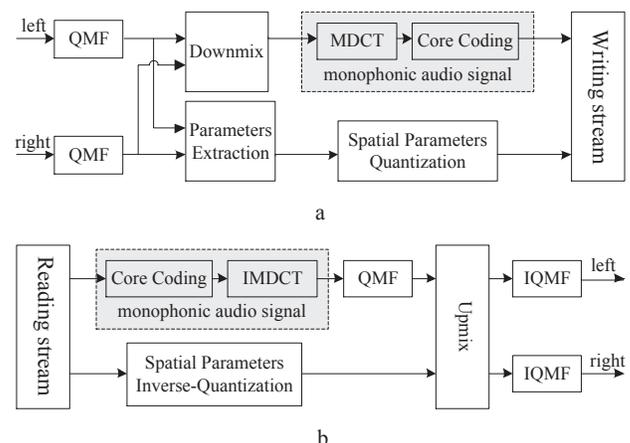


**Fig. 2** Time-domain spatial parameters (a) encoding structure and (b) decoding structure
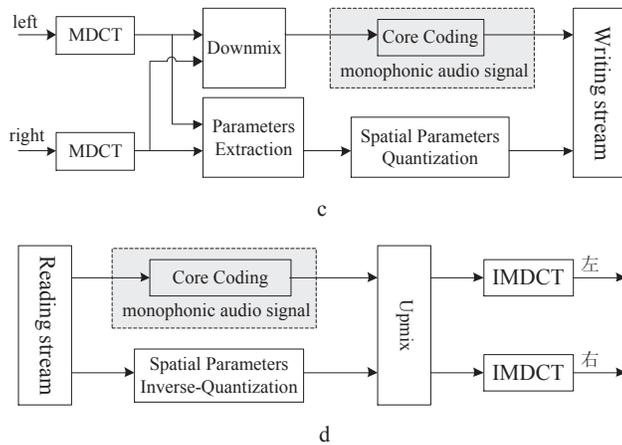
**Fig. 3** MDCT-domain spatial parameters (c) encoding structure and (d) decoding structure

Among them, a mixed module, said about two-channel stereo audio signals are combined into a single channel. On mixed modules On the contrary, it is a single channel audio signal and spatial parameters of signal processing to restore the stereo with the left and right channels. Part of the dashed box marked mono audio codec module, and the contents of the two domains is slightly different. In the framework of the time domain, the dashed box in the original mono codec contain time-frequency transform module and core codec module; in the MDCT-domain framework, the dashed box in the single-channel codec some minor changes.

### 3.2 SPCS ALGORITHM IN MDCT DOMAIN

Here, we first give the MDCT and its inverse transform formula.

$$X(k) = \sum_{n=0}^{N-1} x(n)w(n)\cos[\frac{2\pi}{N}(n+\frac{1}{2}+\frac{N}{4})(k+\frac{1}{2})], \quad (1)$$

$$k = 0,1,2,\ldots,\frac{N}{2}-1$$

$$x(n) = \frac{4}{N} \sum_{k=0}^{\frac{N}{2}-1} X(k)w(n)\cos[\frac{2\pi}{N}(n+\frac{1}{2}+\frac{N}{4})(k+\frac{1}{2})], \quad (2)$$

$$n = 0,1,2,\ldots,N-1$$

Among them, $x(n)$ is time-domain signal, $X(k)$ indicates that the signal of the corresponding MDCT. N is the transform total length. n represents the serial number of time-domain signal, the value of $0 \sim N-1$, k denotes

the MDCT spectrum sequence number value $0 \sim N/2 - 1$.

ILD, ITD and IC three parameters make a detailed description. The ILD parameter frequency domain approach to computing, MDCT-domain ILD formula is defined as:

$$ILD = 10\log_2 \frac{E_{Zl}}{E_{Zr}} \quad (3)$$

$E_{Zl}, E_{Zr}$ respectively, represents left and right channels of energy in the signal under the MDCT-domain. The formula follows the structure of the time-domain the ILD calculation formula, which retains the advantages of the original formula is simple and intuitive.

Numerically, the MDCT transform to transform the structure of the same calculation with DFT, and DFT has a time-frequency signal energy conservation relationship, MDCT spectrum must also contain the time-domain signal energy. The frequency domain method is tentatively scheduled for the above formula, the ILD value when the frequency domain with the derivation of relations can be used for actual coding.

ITD parameter frequency domain approach to computing: ITD parameter is built based on the time-domain analysis, the premise of its existence is to the left and right channel signal has a greater similarity. In the MDCT-domain, the time difference between the concept has changed. The left and right channels of the time difference can not express ITD represented by the parameter space, the ITD value calculations require the adoption of new approaches.

According to the fundamental nature of time-frequency transform, the time domain the time information corresponding to the phase information in the transform domain, the delay size of the time domain frequency domain phase can be calculated. So, the parameters of the ITD in the MDCT-domain a new definition, the group delay of the left and right channels, the formula is as follows:

$$ITD = \frac{d}{dk}\arg(Z_l(k)Z_r^*(k)) \quad (4)$$

Among them, $Z_l, Z_r$ represent the complex spectrum of audio signal, k is the MDCT-domain signal lines serial number. MDCT-domain for solving the signal phase group delay replacing the original time-domain Delay.

However, the MDCT transform is a real number to the transformation of the real number does not meet the above formula the complex spectrum requirements. To build the complex spectrum of the MDCT-domain, the introduction of MDCT transform conjugate MDST transform. The MDST formula as follows:

$$Y(k) = \sum_{n=0}^{N-1} x(n)w(n)\sin[\frac{2\pi}{N}(n+\frac{1}{2}+\frac{N}{4})(k+\frac{1}{2})], \quad (5)$$

$$k = 0,1,2,\ldots,\frac{N}{2}-1$$

$$x(n) = \frac{4}{N} \sum_{k=0}^{\frac{N}{2}-1} Y(k)w(n)\sin[\frac{2\pi}{N}(n+\frac{1}{2}+\frac{N}{4})(k+\frac{1}{2})],$$

(6)

$$n = 0, 1, 2, \ldots, N-1$$

Among them, $x(n)$ represents time domain signal, $Y(k)$ indicates that the MDST domain signal.

Comparison formula, we can only MDST with MDCT transform-based differences, and the cosine transform-based orthogonal relationship exists between. Therefore, they can make a linear combination of a new transform, named MDFT. MDFT and the MDCT and MDST to build the relationship as follows:

$$Z(k) = X(k) + jY(k)$$

(7)

Among them, $Z(k)$ represents the MDFT spectrum, $X(k)$ indicates that MDCT spectrum, and $Y(k)$ indicating that the MDST spectrum. The formula indicates that the the MDFT spectrum is based on the MDCT spectrum of the real part of complex frequency signal to MDST spectrum for the imaginary part.

Transform after the update, the ITD in the frequency domain for solving the formula:

$$ITD = \frac{d}{dk} \arg(Z_l(k)Z_r{}^*(k))$$

(8)

$Z_l, Z_r$ on behalf of the newly constructed left and right channel MDFT spectrum.

IC parameter frequency domain approach to computing:

Domain signal in the MDCT-domain, the signal spectrum is another manifestation of one relationship. When the left and right channel signals in the time domain performance is similar to the frequency domain signal will have similar values, the time-frequency correlation metrics are intrinsically linked. MDCT-domain ITD parameter formula, expressed as:

$$IC = R(Z_l(k), Z_r(k)) = \frac{< Z_l(k), Z_r(k) >}{|Z_l(k)||Z_r(k)|}$$

(9)

Among them, $Z_l, Z_r$ are the left and right channels in the MDCT-domain signal, R indicates the correlation coefficient calculation function Pointed brackets indicate that cross-correlation calculation.

## 3.3 SPCS SYSTEM IMPLEMENTATION

The ultimate purpose of this paper is to achieve a space in the MDCT-domain parameter encoding and decoding the complete system, and has a lower computational complexity and better output quality. During the study period, the paper gives a preliminary system and analyzes its basic performance, the initial completion of the followings :

On the basis of the existing time-domain spatial parameters encoded, modify EaacPlus codec, and parameter encoding and decoding system to complete initial MDCT field space.

Time-domain framework represents the basic process of the EaacPlus codec, its for module replacement can be a new the MDCT field spatial parameters codec, shown in Figure 4: MDCT module position to move and replace the
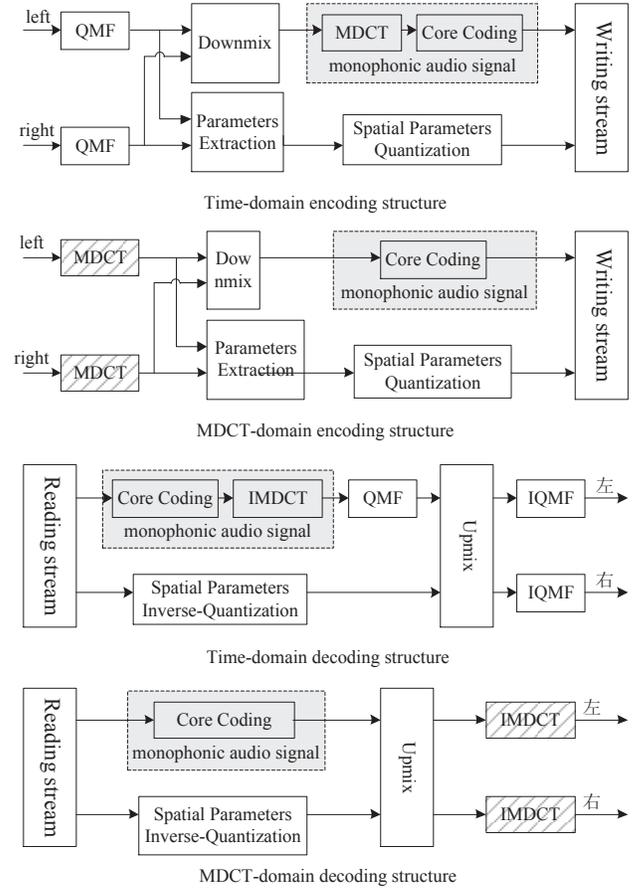


Time-domain encoding structure

MDCT-domain encoding structure

Time-domain decoding structure

MDCT-domain decoding structure

**Fig. 4** Time-domain and MDCT-domain of spatial parameters encoding structure and decoding structure

mono audio coding framework MDCT module out of the encoding side, and replace the QMF filter in front of a mixed module; at the decoder, the mono audio decoding framework the IMDCT module removed and placed on mixed modules. The step function is the audio signal processing transform to the frequency domain, to facilitate the spatial parameters play a role in the frequency domain.

## 4 System Performance Analysis

The system analysis includes two parts of the complex analysis and sound quality. View of the existing system is the intermediate step of the study, so the only system of key indicators of a simple qualitative analysis of specific experimental test is left to the After optimization of each module are given in subsequent articles.

Audio codec system, the common large amount of calculation module MDCT , MDST, QMF, and its inverse transform, etc. Just giving the size of the calculation shown in Table 1. N denotes the frame size, that is, transform window length, it is desirable 256 and 2048, respectively, on behalf of the short window and the window; n represents the number of divided frequency bands in the QMF filter EaacPlus coding n = 64. Due to the MDCT, MDST, IMDCT and IMDST, have the same formula structure, using the same algorithm, their computational complexity is the same. Here, the four used the $N/2$ point FFT fast algorithm.

**Table 1** The module calculates the amount of tables

|  | MDCT | MDST | QMF |
|---|---|---|---|
| Real-Multiplication | $\frac{N}{2}\log_2\frac{N}{4}+2N$ | $\frac{N}{2}\log_2\frac{N}{4}+2N$ | $\frac{N^2}{n}$ |
| Real Addition | $\frac{N}{2}\log_2\frac{N}{4}+N$ | $\frac{N}{2}\log_2\frac{N}{4}+N$ | $\frac{N(N-1)}{n}$ |

The four kinds of transformation are given in Table 1, $N/4$ point FFT fast algorithm, the use of a general algorithm or $N/2$ point FFT fast algorithm of complexity such as shown in Table 2.

**Table 2** The module calculates the amount

|  | Fast Algorithm | $N/2$ FFT | $N/4$ FFT |
|---|---|---|---|
| Real-Multiplication | $\frac{N^2}{2}$ | $N\log_2\frac{N}{2}+N$ | $\frac{N}{2}\log_2\frac{N}{4}+2N$ |
| Real Addition | $\frac{N(N-1)}{2}$ | $N\log_2\frac{N}{2}+\frac{N}{2}$ | $\frac{N}{2}\log_2\frac{N}{4}+N$ |

QMF algorithm is relatively simple, fast algorithm, can be seen from Table 2. However, a large numbers of fast algorithm, such as in the MDCT transform, the computational load significantly reduced, and gradually gained the advantage of low complexity.

Contrast time-frequency large-scale computing the number of modules in the two systems, the statistical list shown in Table 3. It will calculate the amount of each module into the list shown in Table 4.

**Table 3** Frequency of large-scale computing the number of modules in the two systems

|  | Time-domain | | MDCT-domain | |
|---|---|---|---|---|
|  | Encoding | Decoding | Encoding | Decoding |
| Large-scale ComputingModules | 2 QMF 1 MDCT | 1 QMF 2 IQMF 1 IMDCT | 2 MDCT 2 MDST | 2 IMDCT |

**Table 4** Complexity of large-scale calculation modules

| | Time-domain system | |
|---|---|---|
| | Encoding | Decoding |
| Real-Multiplication | $2\times\frac{N^2}{n}+(\frac{N}{2}\log_2\frac{N}{4}+2N)$ | $3\times\frac{N^2}{n}+(\frac{N}{2}\log_2\frac{N}{4}+2N)$ |
| Real Addition | $2\times\frac{N(N-1)}{n}+(\frac{N}{2}\log_2\frac{N}{4}+N)$ | $3\times\frac{N(N-1)}{n}+(\frac{N}{2}\log_2\frac{N}{4}+N)$ |
| | MDCT-domain System | |
| | Encoding | Decoding |
| Real-Multiplication | $4\times(\frac{N}{2}\log_2\frac{N}{4}+2N)$ | $2\times(\frac{N}{2}\log_2\frac{N}{4}+2N)$ |
| Real Addition | $4\times(\frac{N}{2}\log_2\frac{N}{4}+N)$ | $2\times(\frac{N}{2}\log_2\frac{N}{4}+N)$ |

Table 4 calculation to get, as long window (N = 2048) treatment, the time-domain system transformation needs 354,304 times a real number multiplication and 350,048 times the real number addition; 79872 times a real number multiplication and 67,584 times the real number addition to the frequency domain system transformation. As a short window (N = 256) processing, the transformation of time-domain system requires 7680 real multiplications and 7148 real additions; the transformation of frequency-domain system requires 7680 real multiplications and 6144 real additions. The list shown in Table 5. The data show that the

**Table 5** Frequency of large-scale computing the number of modules in the two systems

| | Time-domain System | | | MDCT-domain System | | |
|---|---|---|---|---|---|---|
| N=2048 | Encoding | Decoding | Total | Encoding | Decoding | Total |
| Real-Multiplication | 144384 | 209920 | 354304 | 53248 | 26624 | 79872 |
| Real Addition | 142272 | 207776 | 350048 | 45056 | 22528 | 67584 |
| | Time-domain System | | | MDCT-domain System | | |
| N=256 | Encoding | Decoding | Total | Encoding | Decoding | Total |
| Real-Multiplication | 3328 | 4352 | 7680 | 5120 | 2560 | 7680 |
| Real Addition | 3064 | 4084 | 7148 | 4096 | 2048 | 6144 |

frequency-domain system in the long windows dramatically reduce the computational complexity, thanks to time-frequency Transform Algorithm for the N, whichever is greater, to show the great advantages of. However, in the short window when the complexity of the calculation of frequency-domain system gains little. This

phenomenon stems from the use of more complex modules in the frequency domain system encoder for parameter encoding two times MDST. If the MDST computing good frequency-domain system can handle, you can further reduce the computational complexity of the system, especially in the short window also shows significant computational advantages.

Therefore, in the frequency domain system build is completed, need to focus on solving the computing problems of the MDST to simplify the new entrants calculation module to further reduce the computational complexity of the system, to ensure that the frequency domain of the new system of low complexity advantage.

## 5 CONCLUSIONS

Starting from the significance of the spatial parameters of the MDCT-domain, build a new space parameters of the MDCT-domain codec framework, complete the basic functions of spatial parameters encoded in the MDCT domain. This research makes a contribution to the spatial parametric technology in MDCT domain, and provides a new train of thought to reduce complexity of EaacPlus system.

## Acknowledgement

## References

[1] X.Z.Wen, Research on mask effect of psychoacoustics[D]South China University of Technology (2005).
[2] Christof FALLER, PARAMETRIC CODING OF SPATIAL AUDIO[D]. Lausanne: EPFL (2004).
[3] Frank Baumgarte, Christof Faller. Binaural Cue Coding Part I: Psychoacoustic Fundamentals and Design Principles.[J]. IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESING 11, 509-519 (2003).
[4] Frank Baumgarte and Christof Faller, Binaural Cue Coding Part II: Schemes and Applications[J]. IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESING 11, 520-531 (2003).
[5] J.Herre, K.Brandenburg and D.Lederer, Inrensity Stereo Coding, In Proc.AES 96th conv, preprint 3997, Ansterdam (1994).
[6] Samsudin, Kurniawati Evelyn, Poh Ng Boon, A stereo to mono dowmixing scheme for MPEG-4 parametric stereo encoder[J]. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings 5, 529-532 (2006).
[7] Jeroen Breebaart, Parametric Coding of Stereo Audio[J]. Applied Signal Processing 9, 13051322 (2005).
[8] K.Kuriyama, S.Sano, S.Furuichi, A Precise Estimation of the Computational Complexity in Shor's Factoring Algorithm[J]. Appl. Math. Inf. Sci. 3, 313-322 (2007).
[9] J.S.Ardabili, N.Aghayi, The Modified Interior Point Algorithm for Linear Optimization[J]. Appl. Math. Inf. Sci. 2, 197-205 (2009).
[10] S.D.You, F.Y.Cheng, Spatial Localization Evaluation Model for Parametric Stereo Audio[J]. Appl. Math. Inf. Sci. 6, 397-402 (2012).
[11] L.De, K.JongWeon, Secure Image Forensic Marking Algorithm using 2D Barcode and Off-axis Hologram in DWT-DFRNT Domain[J]. Appl. Math. Inf. Sci. 6, 513-520 (2012).
[12] H.Wei-Chih, L.Cheng-Hsiu, Using Script Command to Conquer the Narrowband Constraint in Synchronous Long-distance Teaching System[J]. Appl. Math. Inf. Sci. 6, 685-695 (2012).
[13] Erik Schuijiers, Jeroen Breebarrt, Heiko Purnhagen, Low complexity parametric stereo coding, In: Audio Engineering Society. Convention Paper 6073[C]. Berlin, Germany (2004).
[14] 3GPP TS 26.405 V6.1.0. 3rd Generation Partnership Project; Technical Specification Group Services and Systerm Aspects; General audio codec audio processing functions; Enhanced aacPlus general audio codec; Encoder specification AAC part(Release 6)[S]. Valbonne, FRANCE (2005).

Appl. Math. Inf. Sci. **7**, No. 2L, 417-423 (2013) / www.naturalspublishing.com/Journals.asp

423

**Cao Sheng** did his BE in Computer Science and Technology and ME in Software Engineering from Computer School of Wuhan University (CSWU), Wuhan, China. Currently he is doing his PhD in Communication and Information System at National Engineering Research Center for Multimedia Software, Wuhan, China. His areas of interest are spatial audio coding, algorithms and multimedia communications.

**Hu Ruimin** has done his PhD in the year 1994 from Huazhong University of Science and Technology(HUST), Wuhan, China . He is currently Professor and Deputy Director at National Engineering Research Center for Multimedia Software. His research interests include computation intelligence and multimedia communications.

**Peng Yuxing** is currently Professor at National University of Defense Technology. His research interests include Cloud Computing and multimedia communications.