

Algorithm for Digital Speech Compression of K-L Transform

Xiaohang LI

College of Computer, Qinghai Normal University, Xining 810008, Qinghai, China

Zhancai JIANG

Physics Department of Qinghai Normal University, Xining 810008, Qinghai, China

ABSTRACT: To realize effectively compression of digital speech, a kind of speech framing compression algorithm based on K-L transform is put forward. Covariance matrix is constructed by speech frame vectors; orthogonal matrix corresponding to frame is got by decomposing eigenvalues of covariance matrix; the transform coefficient vector is got by orthogonal transform of frame vectors with orthogonal matrix, and it is truncated; then frame vectors are reconstructed by truncated transform coefficient vector; finally, the compression of frame vectors is realized by extracting. Simulation experiments show that the algorithm has high data compression rate, reconstituted speech is clear and natural, what's more, good self-adaption denoising of speech is realized.

KEYWORD: K-L transform; Covariance matrix; Orthogonal transform; Data compression

1 INTRODUCTION

As a kind of natural signal, the time-varying property of speech signal makes digital speech sampled actually exist redundancy to a large extent. Therefore, research of digital speech compression technology has become one of hot topics in the field of world communication[1,2]. At present, speech compression algorithm mainly has lossless compression and lossy compression. And lossy compression algorithm can save memory space and reduce digital rate, realizing efficient compression is development direction of current digital speech communication and storage. With the rapid development of computer technology and digital signal processing technology, new digital speech compression techniques are put forward constantly. Compression rate of digital speech compression technology based on wavelet transformation is high, and reconstituted speech is clear and natural, however, the algorithm is relatively complicated[3,4,5]. Speech compression scheme of using Discrete Cosine Transform (DCT) is speech compression of orthogonal transform in essence, the algorithm theory is simple, and has fast algorithm, thus it is adopted widely[6,7], however, inter-frame self-adaption can't be realized. Promoting speech quality after compression processing using the methods of inter-frame self-adaption and intra-frame self-adaption is mentioned in the reference [8], and obvious effects are obtained, but the algorithm is

relatively complex, therefore, it goes against digital speech transmission and storage. Digital signal self-adaption compression algorithm based on speech frame energy classification and inter-frame location inertia is put forward in the reference [9], under the condition of ensuring certain recovery speech quality, the algorithm can reduce recovery time of speech signal evidently, although algorithm complexity is significantly lower compared with reference[8], observed value referred to in the text is an empirical value and can't be obtained easily. In this paper, digital speech compression method based on K-L transform, not only whose algorithm is simple, compression rate is high, reconstructed speech is clear and natural, but also which can make speech self-adaption denoising realize conveniently and effectively because of inter-frame self-adaption.

2 K-L TRANSFORM

K-L transform is short for Karhunen-Love transform, it is a kind of special orthogonal transform, and is mainly used in data compression of one-dimension and two-dimension signal. One wide-stable real random vector $\mathbf{x} = [x(0), x(1), \dots, x(N-1)]^T$, whose covariance matrix \mathbf{C}_x is defined as:

$$\mathbf{C}_x = E\{(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^T\} = \begin{bmatrix} c_{00} & c_{01} & \cdots & c_{0(N-1)} \\ c_{10} & c_{11} & \cdots & c_{1(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ c_{(N-1)0} & c_{(N-1)1} & \cdots & c_{(N-1)(N-1)} \end{bmatrix} \quad (1)$$

In this equation, $E\{\bullet\}$ stands for average operation, $\boldsymbol{\mu}_x = E\{\mathbf{x}\}$ is average vector of signal \mathbf{x} , elements of \mathbf{C}_x are as follows:

$$C_x(i, j) = E\{(x(i) - \mu_x)(x(j) - \mu_x)\} = C_x(j, i)$$

That is to say, covariance matrix is symmetrical. Obviously, matrix \mathbf{C}_x reflects the correlation among each component of signal vector. If each component of \mathbf{x} is uncorrelated, the elements of \mathbf{C}_x except for diagonal are all zero.

The thought of K-L transform is to find orthogonal matrix \mathbf{A} , thus making covariance matrix of transform \mathbf{y} of \mathbf{x} with \mathbf{A} become one diagonal matrix.

According to N order polynomial of λ : $|\lambda\mathbf{I} - \mathbf{C}_x| = 0$, eigenvalues $\lambda_0, \lambda_1, \dots, \lambda_{N-1}$ of covariance matrix \mathbf{C}_x can be obtained, then according to equation:

$$\mathbf{C}\mathbf{A}_i = \lambda_i\mathbf{A}_i, \quad i=0,1,\dots,N-1 \quad (2)$$

N feature vectors $\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{N-1}$ of covariance matrix \mathbf{C}_x can be obtained, and then they are normalized. Normalized orthogonal matrix \mathbf{A} is constructed by normalized vectors $\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{N-1}$, namely, $\mathbf{A} = [\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{N-1}]^T$.

Calculating $\mathbf{y} = \mathbf{A}\mathbf{x}$, and K-L transform of signal \mathbf{x} can be realized.

Reconstructed signal is to solve reverse transformation, it can be completed by following formula:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} = \mathbf{A}^T\mathbf{y} \quad (3)$$

Data compression of $\mathbf{x}(n)$ is conducted, transformation \mathbf{y} of \mathbf{x} is directly compressed, namely, a portion of $\mathbf{y}(n)$ components are gave up. Without loss of generality, supposing that $y(m+1), y(m+2), \dots, y(N-1)$ are gave up, in this case, when $\mathbf{x}(n)$ is recovered by $y(0), y(1), \dots, y(m)$, $\mathbf{x}(n)$ can only be approximated, namely,

$$\hat{\mathbf{x}} = \sum_{i=0}^m y(i)\mathbf{A}_i \quad (4)$$

3 THE PRINCIPLE OF DIGITAL SPEECH COMPRESSION ALGORITHM BASED ON K-L TRANSFORM

Speech signal has short-time stationarity, therefore, speech signal is divided into speech frame $\mathbf{x}(n)$ whose length is 160 sample points, according to equation (1), covariance matrix \mathbf{C}_x is constructed after DC of $\mathbf{x}(n)$ is removed, and orthogonal transform matrix \mathbf{A} is constituted after decomposing eigenvalues, signal \mathbf{y} can be obtained after K-L transform of $\mathbf{x}(n)$ with \mathbf{A} ; and \mathbf{y} is truncated to signal \mathbf{y}^1 whose length is N; according to equation (3), K-L reverse transformation of \mathbf{y}^1 is conducted to reconstitute speech signal of 160 sample points. Reconstructed speech signal of 160 sample points is extracted to N points, then they are transmitted to decode terminal and recovered to speech frame of 160 points by interpolation technology, speech is composited by inter-frame splicing technique. Figure 1 gives schematic diagram of digital speech compression algorithm based on K-L transform.

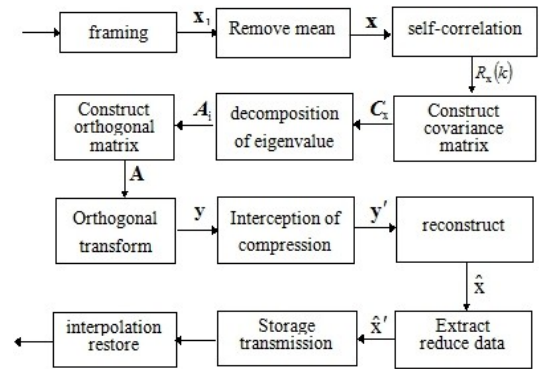


Figure 1. Schematic diagram of digital speech compression based on K-L transform

4 SIMULATION EXPERIMENT OF DIGITAL SPEECH COMPRESSION ALGORITHM BASED ON K-L TRANSFORM

On the basis of figure 1, simulation experiment is conducted using Matlab. In simulation experiment, several speech samples are recorded using the function of sound recorder on the computer: (1) "blue sky, white cloud, green sea", (2) "Leaving the city of White King crowned with rainbow clouds at dawn", (3) "Hoeing up the crops in the sweltering sun at noon, sweat drips onto the soil beneath the crops". All samples are linear PCM codes using 8kHz sampling and 8-bit quantization.

4.1 Comparison between reconstituted speech and original speech

Taking “blue sky, white cloud, green sky” for example to conduct simulation experiment. Figure 2 is 2-times and 4-times reconstructed speech and original speech of digital speech compression based on K-L transform.

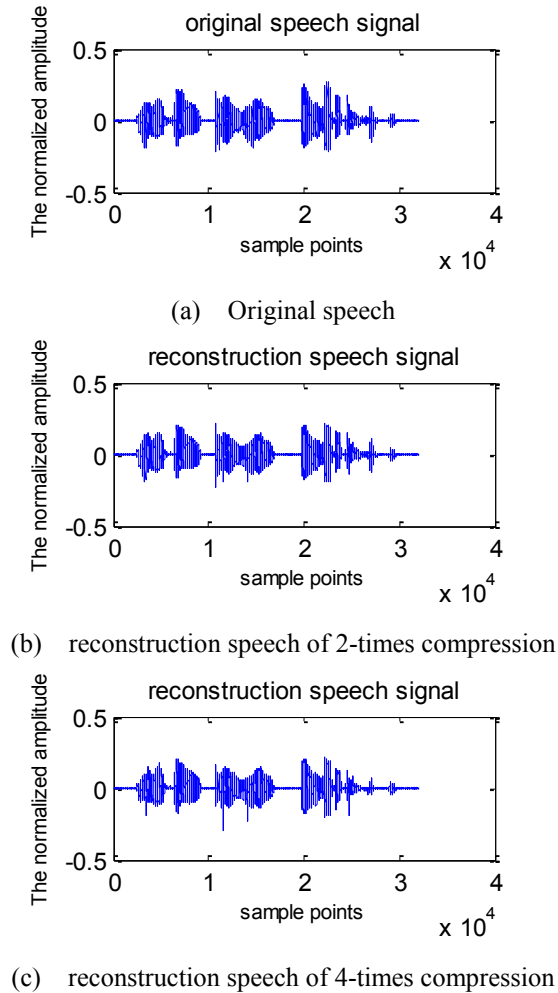


Figure 2. Comparison between reconstruction speech and original speech

Figure 2 shows that compressed reconstruction speech signal waveform conforms to original speech signal waveform approximately. According to audio-visual perception, the intelligibility of reconstruction speech of 4-times compression is 100%, and has good clarity and naturalness.

4.2 Simulation of identical speech sample under different SNR

Speech with noise under SNR of 18dB and 10dB separately, according to the algorithm in this paper, reconstructs speech waveform after 4-times compression.

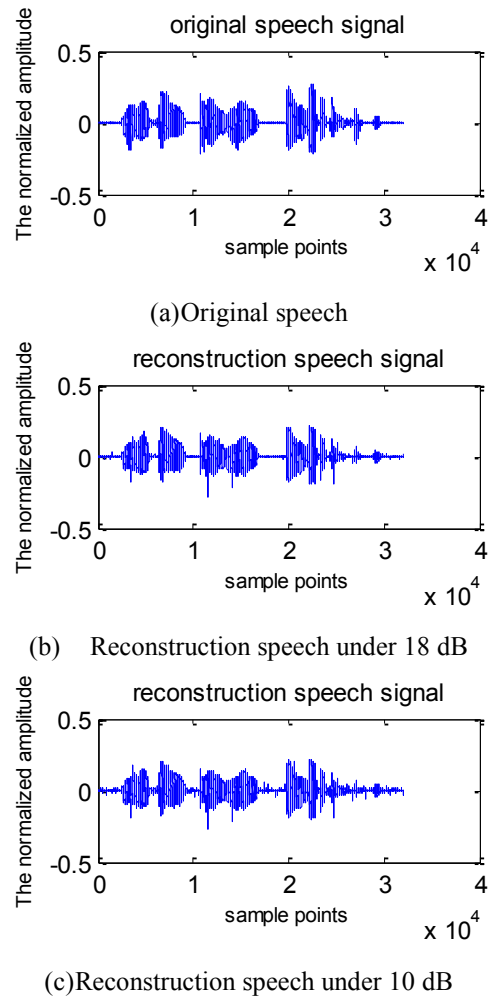
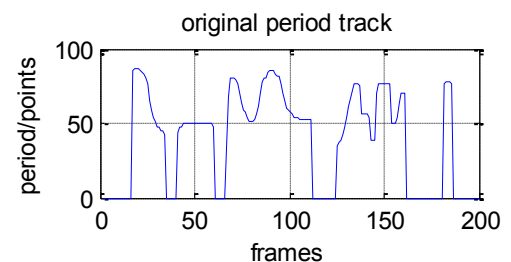


Figure 3. 4-times compression under different SNR

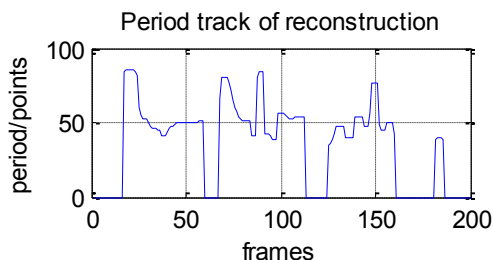
It can be seen from figure 3, reconstruction speech waveform of 4-times compression is approximate to original speech waveform under different SNR, which indicates that not only information loss of reconstruction speech is small, but it also has strong denoising ability to random background noise.

4.3 Comparison of pitch period tracks between reconstruction speech and original speech

Pitch period is one important feature of speech, and is one of main features to distinguish semantic and speakers, many speech processing systems all depend on pitch period. Pitch period tracks of original speech and reconstruction speech after 4-times compression are separately obtained.



(a) Pitch period track of original speech



(b) Pitch period track of reconstruction speech

Figure 4. Comparison of pitch period tracks

From figure 4, the trend of pitch period track of reconstruction speech is approximately consistent with original speech, however, it is not as smooth as original speech, even the period of a small part of frame decreases. This is caused by processing such as truncation, extraction, and it can be improved by selecting appropriate interpolation function.

5 CONCLUSION

Digital speech compression algorithm based on K-L transform is simple, compression rate is high, and reconstruction speech is clear and natural, moreover, both inter-frame self-adaption of speech and self-adaption denoising can be realized. The algorithm is expected to be used for digital speech multimedia storage and digital speech communication.

REFERENCES

- [1] Hu, Guang-shu. 2003. *Digital signal processing(the second edition)*. Beijing: Tsinghua University Press. (in Chinese)
- [2] Yang Zhen-zhen. & Yang Zhen. 2014. *Hard threshold gradient pursuit reconstruction algorithm for speech compressed sensing*. *Signal Processing* 30(4): 391-396. (in Chinese)
- [3] Keon Cha Keon. et al. 2001. *Subband image coding with biorthogonal wavelets*. *IEICE Trans Fundamentals* 75(7): 871-881.
- [4] Hua Rong. 2009. *Speech signal data compression based on non-complete wavelet packet resolution*. *Computer Engineering & Design* 30(10): 2471-2474. (in Chinese)
- [5] Zhou Kai-li. 2002. *Speech signal compression based on wavelet transform*. *Natural Science Journal of Hainan University* 20(2):103. (in Chinese)
- [6] Fu Xiao-juan. & Yang Wan-quan. 2006. *Speech signal compression schemes based on discrete cosine transform*. *Information Technology* (11):74-76. (in Chinese)
- [7] Zhang Chang-qing. & Chen Yan-pu. 2014. *Discrete cosine wavelet packet transform and compressed sensing for speech signal*. *Technical Acoustics* 33(1): 36-40. (in Chinese)
- [8] Guo Hai-yan. et al. 2010. *Adaptive speech compressed sensing in the DCT domain*. *Chinese Journal of Scientific Instrument* 31(6):1262-1268. (in Chinese)
- [9] Lei Ying. et al. 2012. *Adaptive inter-frame speech compressed sensing*. *Signal Processing* 28(6): 895-897. (in Chinese)